UNIVERSITY OF PENNSYLVANIA

THE COMPETENCY OF FIFTY COLLEGE STUDENTS

(A DIAGNOSTIC STUDY)

BY

KARL GREENWOOD MILLER

A THESIS

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

PHILADELPHIA

1922

UNIVERSITY OF PENNSYLVANIA

THE COMPETENCY OF FIFTY COLLEGE STUDENTS

(A DIAGNOSTIC STUDY)

BY

KARL GREENWOOD MILLER

R. UNIV. BIBLIOTHEEK LEIDEN

A THESIS

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

PHILADELPHIA

1922

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547328_0037

THE COMPETENCY OF FIFTY COLLEGE STUDENTS.

(A Diagnostic Study.)

NOTE

This Thesis will be found reprinted as No. VIII of EXPERIMENTAL STUDIES IN PSYCHOLOGY AND PEDAGOGY

INTRODUCTION.

No task more worthy of attention confronts the psychologist today than the scientific study of the college student by means of mental tests.

Psychological tests were first employed in the examination and segregation of the mentally feeble. A large number of clinics connected with modern school systems, hospitals, or juvenile courts have found these tests of service in detecting mental subnormality. It has only been in the last decade, however, that the possibilities of the psychological examination of "normal" individuals have been recognized, and rapid advances are now being made in this field. The success with which mental tests were used in the classification and stratification of the great mass of men who formed our National Army probably did more to bring about a general acceptance of the method and principles involved than would have resulted from many years of experimentation in peace times. Today, psychological tests are used not only in the field of education but also form an integral part of the selective and administrative machinery of many large industrial organizations. The present vogue of the mental test carries with it one real danger in that the uninitiated are likely to demand more of the psychologist than he can give.

Without doubt it is now possible to say, as a result of a psychological examination, that one individual possesses too little mentality to admit of his being a self-supporting member of society, that another can be trained to perform a simple task satisfactorily, that a third has ability which will enable him to fill a place in the great middle class, while still another has intellectual endowments which should lead him into the fields of higher education and professional activity. These broad classifications can be made through the employment of many and various tests which have been carefully devised and scientifically standardized. With the concept of differing levels of general intelligence fairly well developed the psychologist now faces the task of classifying individuals. When the attempt is made not only to ascertain the general performance level but also to determine for what occupation the specific abilities of the individual best fits him, the difficulty of the problem is tremendously increased. Shall the man of small competency be a ditch-digger or a stevedore?

Is the citizen of mediocre ability best qualified to follow the vocation of motorman, mechanic or clerk? Should the college student be guided into industry, law or teaching?

These questions imply that the psychologist must also function as a vocational adviser, and while this obligation may not at present be generally accepted, the implication is nevertheless warranted. Mental tests, if they are to be of value to society, must lead to prognoses as well as to diagnoses and must at least offer to the individual tested some information which may be useful in the attainment of greater personal and social efficiency. In much the same manner as the employment manager of today places the applicant in some particular position in his organization, so the psychologist of the future may find it possible to direct each member of society to the one vocation which will best utilize his peculiar qualifications.

It is hardly necessary to point out that the problem of differentiation becomes increasingly complex as the higher levels of intellectual organization are approached. The idiot may be consigned to custodial care with but small probability of error. The stevedore, the scavenger, and the ditch-digger gravitate to their respective occupations without perceptible friction. The "common people" present a more difficult problem in view of their higher level of performance and greater complexity of response, but even here noteworthy advances have been made in recent years through the introduction of vocational guidance and the application of psychological principles to industrial management. Although investigation of this character has hardly passed beyond the experimental stage, a beginning has nevertheless been made, and remarkable developments during the next decade may be confidently anticipated.

The task of differentiating the particular abilities required of the successful plumber, mechanic, clerk, motorman, and telephone operator—to mention only a few of the almost countless range of occupations—is doubtless a difficult one, but it hardly approaches the complexity of the problem presented in the guidance of individuals of greater intelligence and higher intellectual organization to the one vocation for which each is best fitted. While interest, personality, and various external circumstances can not be disregarded as important factors in the selection of the life work, the concern of the psychologist lies primarily in the determination of the specific abilities requisite to each type of professional activity, and in the scientific evaluation of the particular abilities possessed by each individual. It is with the latter phase of the problem that this investigation will deal, the interest being centered on the college student, who, despite his many shortcomings, must be regarded as representative of the highest intellectual type of young manhood in the country.

Historical.

The attempt to appraise the undergraduate by means of mental tests must not be considered a new departure in the field of psychology. The credit for the first scientific study of the American college student goes to J. McKean Cattell. Stimulated by his researches in the anthropometric laboratory of Francis Galton, he inaugurated in 1887 a series of experiments with undergraduates at Harvard University, which investigation he continued at the University of Pennsylvania and Bryn Mawr College in 1888 and 1889, and in the following years at Columbia University. The report entitled, "Physical and Mental Measurements of the Students of Columbia University", which appeared in the *Psychological Review* for November, 1896, and in which Professor Cattell collaborated with Dr. Livingston Farrand, was probably the first publication of the results of a systematic study of the mental status of the college student. This report is of peculiar interest today not only because of its scope, but also in view of the surprising number of mental and physical tests actually employed or suggested at that time which now constitute the accepted instruments of every clinical psychologist. While the purpose of the investigation was necessarily the establishment of norms by the statistical treatment of the test results of one hundred students, and the aim of the present study is rather the observation of individual variation, it will nevertheless be of interest to indicate briefly the character of the information recorded by Cattell. Anthropometric measurements such as height, weight, and cephalic diameters were noted, and in addition such physiognomic characters as the color of hair and eyes, and the size and shape of ears. In addition, psychophysical determinations of visual and auditory acuity, sensitivity to pain, and various types of reaction time were made, as well as tests of a more strictly psychological nature which included memory of drawn lines, memory of numbers heard, cancellation test, color preference, types of imagery, and others.

The investigation under consideration was carried on during the academic years of 1894–95 and 1895–96, and it is of interest to note that the results were published so as to be of assistance to a committee appointed at the annual meeting of the American Psychological Association held at Philadelphia in December, 1895, to consider the feasibility of co-operation among the various psychological laboratories in the collection of mental and physical statistics. This "Committee on Mental and Physical Tests", which consisted of Professors Cattell, Baldwin, Jastrow, Sanford, and Witmer, may well be said to have laid the foundation for all subsequent develop-

ments in the realm of psychological tests in its report to the Psychological Association at the meeting held in Boston in 1896. This report may be found in the *Psychological Review* of March of the following year.

Having thus briefly indicated the inception of the present field of investigation, it would be a thankless task to trace its history down to the present moment in any adequate manner. Studies of this character have been carried on in every psychological laboratory connected with a college or university, and a complete bibliography of the reports on the subject would cover many pages. It will be well, however, to mention a few of the more important investigations which have a direct bearing on the present problem, in so far as it concerns the correlation of test results with academic standing. Wissler (1) correlated the results published by Cattell and Farrand, to which reference has been made above, with the university grades assigned to the hundred students under consideration. Calfee (2) has reported on "Four General Intelligence Tests" given to approximately one hundred students at the University of Texas. Similar investigations have been made by Rowland and Lowden (3) at Reed College, Waugh (4) at Beloit College, and by Kitson (5) at the University of Chicago. The latter study is particularly worthy of note in that a very careful and intensive examination of forty students was made. King and McCrory (6) report the results of tests on five hundred freshmen at the University of Iowa, Caldwell (7) has correlated the Intelligence Quotient of approximately one hundred students at Randolph-Macon Woman's College, as determined by the Adult Tests of the Stanford Revision, with college grades, and Rogers (8) gives interesting results of her investigation at Goucher College. In the reports mentioned above, Kitson and Caldwell also record correlations between test results and estimated intelligence, which will be referred to later in this discussion. Incomplete as is the preceding sketch, it nevertheless gives some indication of the wide-spread interest in the application of mental tests to the college student. In this connection it will likewise be well to refer to the comparatively recent development in the field of psychological entrance examinations, which are now demonstrating their practicability in a number of the larger universities, and which constitute a further ramification of the same problem.

Experimental Conditions.

Stated briefly, the aim of the present study is to examine certain data which have been collected relative to each member of the class in elementary psychology at the University of Pennsylvania during

the academic year 1919–20. This information consists of the score obtained in a "general intelligence examination", the results of a series of psychological tests, a rating on estimated competency, and a rating based on the academic standing of the individual as determined by the final grades received in all courses completed at the University. The treatment of results will be concerned with the examination of correlations existing between the various ratings under consideration, and with the scrutiny of the individual record with a view to reaching, if possible, some conclusions which might be of assistance to the student in the direction of his intellectual development.

The investigation differs from many which have preceded it, in that the psychological tests, with one exception, were given as a part of the ordinary class instruction and therefore not primarily as tests. The elementary work in psychology consists of two courses known as Psychology 1 and 2, each requiring five hours of class attendance and continuing throughout one semester. Since credit in Psychology 1 is prerequisite to admission into Psychology 2, the two courses may be considered as a single introductory course lasting through the full academic year. Of the five hours of class attendance per week, only one hour is occupied by a formal lecture, the remaining four hours being devoted to laboratory work. During the first semester a number of mental tests are given as a part of the laboratory work and with the purpose of graphically demonstrating the various factors which function in the formation and development of the intellect. It is believed that this method enables the student better to understand and appreciate the particular ability or mental process under discussion. It is not claimed, therefore, that the series of tests employed would necessarily have been chosen had the purpose been the psychological examination and diagnosis of the individual to the exclusion of other considerations. However, the tests unquestionably provide a very satisfactory framework upon which to build a logical presentation of systematic psychology as well as offering a medium for the demonstration of fundamental psychological processes. In addition, the tests are extremely valuable to the student, in that they enable him to determine his peculiar mental assets and liabilities through a comparison of his individual results with accepted standards or class distributions.

Since the tests under consideration were given as a part of the usual classroom procedure, the scientifically controlled conditions which are generally regarded as indispensable to a psychological investigation of this character were for the most part lacking. As the class in Psychology 1 numbered more than two hundred students,

the laboratory work was conducted in three sections with an average enrollment of approximately seventy. These three sections all met in the same room, one being held at eight-thirty o'clock in the morning, another at two in the afternoon, and the third at three o'clock on a different afternoon. While the time of meeting was constant for each section, the variation in hour possibly affected the comparability of section results. With such a large number of students in a laboratory class, some were necessarily seated at a greater distance from the instructor than others and in addition a few were near windows which may have provided distraction of one kind or another. In some cases, the same test was given to the three sections by different experimenters, and although the attempt was made to adhere as closely as conditions would permit to the standard procedure, this variant may have affected the results to some extent. In summary, lack of uniformity in the time of meeting of the different sections, in the seating arrangement of the classroom, and in the identity of the experimenter may be considered factors which expose this investigation to criticism as being unscientifically conceived and prosecuted.

The comparative absence of controlled experimental conditions, however, cannot be said to invalidate the results. It is an open question whether the environment imposed upon a subject by scientifically controlled conditions elicits a more representative sample of behavior than that produced under less artificial circumstances. Is the psychologist more interested in the reaction of a subject who has been isolated in a sound-proof cabinet with a screen before his eyes to eliminate distracting visual stimuli, or in the behavior of the same individual as displayed in natural association with his fellows? For some, the classroom would provide as unnatural an environment as any that the experimentalist might impose, but for a group of university students no more satisfactory and less distracting atmosphere could be selected than that of the recitation hall or laboratory. It is contended, therefore, that the experimental results here presented provide an index of the mental status of the college student as reliable as any that might have been obtained under other conditions.

Having thus disposed in a somewhat arbitrary manner of any criticisms which might be voiced against the general procedure followed in this investigation, it will be well to consider the treatment of the data collected before undertaking a description of the specific tests employed. As has been indicated, the information available concerning each member of the group here studied consists chiefly of the results of a series of mental tests and the academic record of

the student as displayed in his college grades. The problem of devising some statistical method by which the various scores and grades may be made easily comparable is immediately encountered. For example, a member of the class might have obtained a score of 131 in the general intelligence examination, a time rating of forty-three seconds in a mechanical test, and he may have an audito-graphic memory span of eight digits as well as a number of other test results. In addition, his college record may show that he has received the highest grade in 10 per cent of his academic work, a passing grade in 70 per cent, and that he failed in the remainder of his courses. The necessity of reducing these various values to some common denominator so as to render them comparable is evident.

Perhaps the most natural procedure would have been to obtain arithmetical averages of the results of each test and rate the individual performance in terms of its variation from the average. After determining a rank order in academic standing it would then have been possible to calculate the correlations and intercorrelations desired. Such a method is valuable in the examination and standardization of tests, but it has little to offer when the interest is chiefly centered in the study of the individual rather than the tests, and it has a tendency to obscure significant personal variations under a mass of figures. Indeed, it is probable that correlation as a statistical method has been carried to extremes in recent psychological investigations. When the results of two mental tests show a high degree of correlation, it does not necessarily follow that they tap two abilities which are mutually dependent, but rather that the tests have called the same ability or group of abilities into play. Conversely, a lack of significant correlation may show either that one of the tests is unreliable or that the results are not dependent on some common factor. If college psychological tests are designed to call into play the same abilities which function in college grades, such tests are useless unless a high degree of correlation with academic standing can be demonstrated. On the other hand, the absence of such correlation does not show the tests to be devoid of significance, but merely that they measure other abilities or factors than are predominant in the attainment of grades. Further, if it be admitted that individual competency is the algebraic sum of the various specific abilities and disabilities, then the ideal series of psychological tests—which would include a different test for each special ability—would show no significant intercorrelations for individuals at the same level of general intelligence.

The purpose here, therefore, is to present the material in such form as best to facilitate the scrutiny of the individual record, rather

than in the form most convenient for statistical treatment. Hence the various results must be rated on some common scale which has steps of sufficient number to provide the necessary differentiation without introducing a false accuracy. In addition, since many of the tests used have not been scientifically standardized, it is important to adopt a rating system which will permit the comparison of test scores with each other rather than with accepted standards.

A consideration of the many rating scales which lend themselves to the present purpose shows that the extremes are to be found in the percentile and the two-division systems. It is hardly necessary to enter into a discussion of the pseudo-accuracy of the percentile grade. It is only in the very unusual case that the material to be rated can be clearly enough differentiated to give any real significance to each of the hundred points on the percentile scale. Investigations have shown the wide variation in grades given by different instructors to the same examination paper even in the field of mathematics where the greatest accuracy might be expected. This variation, however, is no greater than that shown in the grades given by the same scorer to the same paper at different times. The injustice done to the college student who receives a final mark of 69 per cent in the course which demands 70 per cent as a passing grade has been commented upon too frequently to require more than passing mention in this discussion. Obviously, the refinement of the percentile scale is too great for the material here at hand. On the contrary, the system which merely distinguishes the "passing" from the "not passing" does not provide sufficient differentiation for analytic examination of the results of a series of mental tests.

Popular acceptance would seem to have stamped its seal of approval on a five-division rating scale. Cabbages and kings alike are usually judged mediocre, good or very good, poor or very poor. The great majority of our quantitative expressions are given in these terms, and the system seems to provide a sufficient number of significant levels without introducing the fallacy of too great refinement. This psychological justification of the five-point scale, as well as other considerations of convenience and facility of comparison, led to its adoption as the most satisfactory method of treating the various results and scores herein presented. In accordance with this decision, the results of each test given to the two hundred students who comprised the class in elementary psychology were arranged in rank order and separated into quintiles. While the nature of some of the tests has made even such a coarse rating as this quite difficult, it is believed that the system adopted is the most practicable that could have been devised for the present purpose. Since all grades

assigned in the School of Arts and Science at the University of Pennsylvania are recorded in terms of a five-point system, an added advantage is gained in the comparison of test scores with academic success.

The results tabulated in a later section will therefore not be found to contain the number of digits for the memory span, the number of seconds required for the completion of the cylinder test, or the number of problems correctly solved in the general intelligence examination, but instead the translation of each of these scores into a quintile rating. If the performance of an individual places him in the best twenty per cent of the class in a particular test, he is given a rating of "5", if in the poorest fifth of the group of two hundred, his quintile grade would be "1". The upper, middle, and lower quintiles are represented by "4", "3", and "2", respectively. By thus evaluating a given performance in terms of the class results, it will be found a relatively simple matter to scrutinize the ratings for each individual and gain a fairly trustworthy impression of his standing in an unselected group of university students, and at the same time to note his peculiar mental assets and liabilities.

Selection of Group.

Since it is the aim of this investigation to discover individual differences in a comparatively homogeneous group of students, it seemed advisable to make certain eliminations before undertaking an intensive study of test scores and college grades. Of the 220 students who registered for Psychology 1 at the beginning of the session of 1919–20, fifteen withdrew before the work of the semester was really under way, reducing the class to an actual enrolment of 205. Of these, 125 were taking the course in the School of Arts and Science, the remainder being students in the School of Education. This split also gives the approximate ratio of men to women in the class. During the semester twenty members of the class were dropped because of deficiency or received a failure upon the termination of the course which excluded them from participation in Psychology 2. Since it was deemed advisable to make the completion of both courses one of the requisites for inclusion in this study, these twenty students were automatically eliminated. In order to obtain homogeneity it was also decided not to introduce sex differences but to limit the investigation to male students enrolled in the School of Arts and Science. Of the 125 men who originally started the course only 113 were eligible for Psychology 2, and of these only eighty received final grades at the end of the second semester. Since one of the ratings to be taken into consideration is based on academic standing,

it was thought best not to include first-year students in the selected group, thereby eliminating all who were not able to survive at least one year of university work, reducing the variation in age, and at the same time making it possible to base the academic rating on college grades received during two or more years of class attendance.

When these eliminations had been made, fifty-one students were eligible for inclusion in this study. Of this number, one individual over thirty years of age was arbitrarily excluded as not conforming to the normal college age. Of the fifty remaining as subjects of this investigation, thirty-three had sophomore standing, twelve were rated as juniors, and five were seniors. The average age for the group as of October 1, 1919, was 20.8 years, that of the sophomores being 20.5 years, and of the juniors and seniors, 21.3 and 21.4 years respectively. Although the averages in the latter cases are not of great significance due to the small size of the groups in question, the figures quoted do show that the larger group of fifty is composed of students of approximately normal college age. In conclusion, it will be well to point out that although only about one-fourth of the total class in psychology is to be included in the study, the selection was made on the basis of group qualifications and without regard to individual merit, except for the automatic elimination of those members of the class who were excluded for deficiency in scholarship.

The Psychological Tests.

The psychological tests included a general intelligence examination, the "Psychological Examination for College Freshmen and High School Seniors", devised by Professor L. L. Thurstone, and the following thirteen tests designed to exercise some particular ability or group of abilities: (1) Ausfrage (Observation) Test, (2) Taylor Number Test, (3) Memory Span for Digits, (4) Memory Span for Syllables, (5) Memory Span for Ideas, (6) Description of Formboard, (7) Trabue Language Test, (8) Courtis Arithmetic Test, (9) Differences and Likenesses Test, (10) Opposites Test, (11) Definitions Test, (12) Humpstone Memory Test, (13) Witmer Cylinder Test.

The tests were given in the order indicated, and, with the exception of the Witmer cylinder test, all were given during the first half of the academic year, or, in other words, as part of the laboratory work in Psychology 1. The cylinder test was given in connection with the competency rating toward the close of the second semester, and it is the only one of the series which was given as an individual and not as a group test, and likewise it alone was given primarily as a test and not for its didactic or illustrative value. Of the series employed, the memory span for digits, the Trabue sentence com-

pletion, the Courtis arithmetic and Witmer cylinder tests are all in general use and have been carefully standardized. The Ausfrage, memory span for syllables and for ideas, description, differences and likenesses, opposites, and definitions tests have merely been adapted to the present instructional aims, while the Taylor number test and the Humpstone memory test are here described for the first time.

Before undertaking a description of the various tests it will be well to note in connection with the scoring that the quintile ratings were in each case based on the results of the class of approximately two hundred students and not on the relative performance of the fifty here to be considered.

Thurstone Psychological Examination.

On the afternoon of October 26, 1919, some fifteen hundred first-year students in the various undergraduate schools of the University of Pennsylvania were given the Thurstone "Psychological Examination for College Freshmen and High School Seniors", the experiment being conducted by the Department of Admissions in co-operation with a number of other colleges and universities in the state of Pennsylvania. At the same hour, the examination was given to approximately 120 students who were then meeting in different sections of Psychology 1, with the purpose of comparing the scores obtained by this relatively selected group, which included no freshmen, with the results of the larger first year group. The fifty students who form the basis for this investigation all took the examination as members of laboratory classes in psychology.

Description: The form which was used is known as "Test IV, Edition of September, 1919—issued by L. L. Thurstone of the Carnegie Institute of Technology". The examination consists of 168 short problems which are to be solved in order. The printed directions on the cover of the pamphlet, and the specific nature of the instructions for each problem greatly simplify the administration of the test. The important timing element, which is a complicating factor in such examinations as the Army Alpha and the Otis intelligence test, is practically eliminated in this case. The directions, which are read by the examinee before the beginning of the examination, state that thirty minutes will be given in which to solve as many problems as possible. The problems are to be taken in order, but instructions are also given to skip any which may not be understood. The task of the examiner, therefore, is merely to call attention to the directions after the pamphlets have been distributed, and to give the appropriate signals at the beginning and end of the thirty-minute period. Although the subject is directed to solve the prob-

lems in order, the final score is determined solely by the number of correct solutions without reference to errors or omissions.

The 168 problems which compose the examination are arranged in what is known as the cycle-omnibus form. In other words, while only six different tests are employed, the separate problems which go to make up each test appear in rotation instead of being grouped together as is more usually the case. The examination may readily be analyzed into a number of sets of eight problems each, and in each set all of the six types of tests occur in regular order. The first two problems in each group form part of a *general information test*, while the next two are a variation of the familiar *analogies test*, and the fifth is a *sentence completion test* taken from the language scales devised by Trabue. The sixth problem in each set is of the type known as the *syllogism test*, and the seventh, referred to by Thurstone as the *reading test*, is a form of the widely-used proverbs test. The last problem of the group is an example of the *number completion test*.

Since eight of the 168 problems are preliminary samples for which the correct solution is given, the examination actually consists of only 160 problems of which forty comprise a test of general information, an equal number form an analogies test, while each of the other types is represented by twenty problems. The final score is therefore weighted in the direction of information and analogies.

Discussion: It is not the present intention to enter into a lengthy criticism of the validity of general intelligence tests. Ever since the Binet-Simon scale came into popular use, this question has been discussed with varying degrees of fervor, and the many recent additions to the store of group tests, which have appeared as an aftermath of the army series, have served to keep the controversy before the psychological eye. Even the most conservative introspectionist must admit that the army tests performed a valuable service in the stratification of the National Army, and that satisfactory results are being obtained at several of the larger universities by the admission of students on the basis of group psychological examinations in lieu of the traditional entrance requirements. The general intelligence test is of established significance in the differentiation of the various well-recognized levels of performance. The question which must be broached here is whether it is of equal significance when applied to individuals at the same general intellectual level, and particularly whether it discloses any information of value relative to the college student.

It may be contended that the Thurstone examination is designed for the elimination of applicants for admission, and that significant

results are not to be expected when the test is applied to students who have not only met the entrance requirements but have successfully completed at least one year of college work, as is the case with the present group. Nevertheless, it seems profitable to inquire into the particular abilities called into play when the examination is submitted to college students. A mere inspection of the series of problems quoted above will demonstrate that the correct solutions could be given by any person of the intellectual level of the college student were unlimited time at his disposal. An exception to this statement must be made in the case of a few general information questions, which are so designed that no individual would be likely to give correct answers to all. Hence, whatever the abilities involved in the solution of the six different tests of which the examination is composed, the score obtained is primarily an index of mental alertness or of the rapidity of the reasoning processes and not of what is usually termed general intelligence. If the colleges wish to admit candidates on the basis of the speed with which a problem can be solved and without regard to the proportion of correct solutions, then the Thurstone examination should be found very satisfactory. Or if experimentation can demonstrate that the rapid thinker is also the accurate thinker, this type of test will be equally acceptable. In this connection it is interesting to note that a correlation between the score on the Thurstone test and the percentage of correct answers to the total number attempted shows the unexpectedly high coefficient of +0.74 (Pearson) in the case of fifty results chosen at random. This would seem to indicate that accuracy and speed are closely related, and must be considered as arguing for the validity of the examination. A study of the same fifty cases shows that on the average only 85 per cent of the solutions given were correct, the syllogism test being the most difficult with 23 per cent incorrect, while the greatest accuracy was shown in the analogies, sentence completion, and number completion tests, each of which had an error of only 10 per cent.

Although the "cycle-omnibus" type of examination has marked advantages, chief among which are simplicity in administration and scoring, one important weakness must be noted. Assuming that the six tests which compose the examination call into play different abilities, it is often desirable to analyze a given score in order to determine individual assets or deficiencies. In other words, a low score might be due either to a poor performance in all six of the tests, or to a particularly deficient result in any one of them, such as the general information test. While the score would be the same in both cases, its significance would be very different. Most of the

general intelligence tests are so arranged that the scores for the different parts of the examination are readily available for comparison. In the case of the cycle-omnibus, however, an analysis of the various test results is practically impossible in view of the undue expenditure of time and effort required.

As in the case of the other tests, the class scores for the Thurstone examination were arranged in rank order and quintiled. The rating for each individual in the table of results shows the quintile grade and not the actual score. A discussion of the results and correlations obtained will appear in a later section.

Ausfrage Test.

Description: This test is a variation of the familiar Ausfrage test, differing from it only in that specific questions are asked. In the first part of the test a picture was thrown on the screen with the aid of a stereopticon and the class allowed to examine it for two minutes, the following instructions having previously been given: "I am going to throw a picture on the screen. While it is there I want you to do nothing but look at it. When I have finished I will ask you to answer some questions." Upon the removal of the picture, ten questions were asked relative to different objects which may or may not have appeared in the picture.

The second part of the test consisted of a series of ten questions based on observation of the university buildings and campus and of the city of Philadelphia. In both parts of the test written answers were obtained.

In scoring the results, each correct answer received one point, giving a maximum score of twenty points. The class results were distributed in rank order and quintile ratings determined.

Discussion: The ability primarily involved in this test is that of observation, which implies attending to something and making note of it for a purpose. In this case, the stimulus was visual, and therefore visual sensibility and discrimination are essential. It may be assumed, however, in connection with this test as well as those which follow, that every member of the class was equipped with the necessary sensibility and with the psycho-motor apparatus involved in the recording of results, and these factors will therefore be disregarded in discussing the various tests. Analytic concentration and distribution of attention play a part in the process of observation, as does the factor of associability, which will be discussed at some length in connection with memory span. Memory enters but little into the first part of the test, since the retention required is of brief duration, but it must be considered an important element in

the second part. While all of the abilities mentioned are involved, the test may be regarded as primarily one of observation.

Taylor Number Test.

Description: The test material consists of a sheet of white paper 8½ x 10 inches in size, upon which are distributed in a haphazard arrangement the numbers from 1 to 50, inclusive, printed in half-inch bold-face black type. One sheet was handed to each student with the numbered side of the paper downward, while the following directions were given: "I am going to give each of you a sheet of paper. I want you to let it lie on your desk until I tell you what to do with it. When I am ready I shall give three commands, the first, 'Ready', the second, 'Turn', and the third, 'Go'. When you turn the paper, turn it from the right side over to the left, and in the upper left hand corner you will find the number '1'. On the paper are the numbers from 1 to 50, not arranged in any regular order, but scattered over the sheet. As soon as you have turned the paper, place your pencil on number 1. When I say 'Go', draw a straight line to number 2, then to 3, and go on in order to each number until I say 'Stop'. When I say 'Stop' hold your pencils in the air immediately."

A time limit of forty seconds was allowed for the test, and the results were then scored on the basis of the highest number reached. The distribution of class results was made and the quintile ratings obtained.

Discussion: In so far as is known, this test was devised by Mr. Charles K. Taylor and was first used a number of years ago in the Psychological Laboratory of the University of Pennsylvania. When repeated a number of times the Taylor number test serves as an excellent index of trainability, but when only one trial is allowed it must be considered a test of alertness or distribution of attention. In many ways this test is similar to the more familiar "Cancellation Test", but it has the advantage of providing no definite cues to exploitation, since great care was taken not to arrange the numbers on the sheet in an orderly manner. In addition, the goal is constantly changing in this test while it remains constant in the cancellation test, where the aim is to locate some particular letter or digit.

It would seem unlikely that discrimination of form would be a factor worthy of consideration in the performance of this test by college students, but under the conditions of rapid exploration which usually exist this element cannot be overlooked. The test also has an important motor phase, and coördination and control of move-

ment play a rather important part in the result. However, the higher scores may be attributed to good distribution of attention coupled with methodical exploration.

Memory Span for Digits.

Description: The material for this test consists of twenty series of digits, ranging in length from three to twelve digits, and including two series of each length. The series used were employed by H. J. Humpstone in his standardization of the test, and were so prepared that no two digits occur in the natural order or in the reversed order, no two succeeding series begin with the same digit, and no digit is repeated except in the series of ten or more. Zero is not used.

The instructions given were those used by Humpstone (9). "This is an experiment. In every experiment it is necessary for everyone who takes part to do just what the experimenter asks. Please do just as I ask you to. I am going to say some numbers. While I say them I do not want you to do anything except look at me and hold your pencils up where I can see them. When I put my pencil down you write on your paper the numbers I have said." The digits were then pronounced at the rate of one per second, without rhythm or change of intonation except that on the last one of a series the voice was allowed to fall as a signal for reproduction. In each case the number of digits in the series was announced before the series was given. In scoring the results, the number of digits in the longest series correctly reproduced is considered the memory span. The quintile ratings are based on the scores thus obtained.

Discussion: According to Professor Humpstone, "It has been assumed by almost everybody who has written on the test that it tests memory. A careful analysis causes us to doubt the validity of the assumption. Some imagination is required. The subject must have enough imageability to get perceptions of the stimuli . . . . In the same sense memory is involved. The images must be retained long enough for reproduction. But this period is so brief that the results do not furnish any criterion by which to judge of retentiveness . . . . Attention is involved also . . . . The ability to distribute the attention well is doubtless an aid in the performance." He continues, "Perhaps the memory span test comes nearer to testing one definite ability than any other test. Whatever other factors or abilities enter into the performance of this test it is clear that the thing specifically tested is the ability to grasp and associate a number of discrete units of perception in a definite order. This is not memory as pointed out above. We are using the term associability and subsuming it under the general heading imagination. Associability

refers to the 'number of discrete perceptions associated in a single act of attention, and the combination of the associated component parts of a single perception'."

While the memory span test is of great value in the examination of the mentally retarded, and it can be said without fear of contradiction that a memory span of four and probably of five is prerequisite to intellectual development, the test loses much of its significance when applied to a group of college students. Certainly in the case of the higher scores the result has been exaggerated by means of grouping, and the factor of planfulness plays an important part. The lower memory spans of five and six digits are probably of greater significance.

Memory Span for Syllables.

Description: The subject-matter of the test consists of sixteen sentences ranging in length from ten to fifty syllables. The series provides two sentences at each of the various levels, namely ten, twenty, twenty-five, thirty, thirty-five, forty, forty-five, and fifty syllables. The sentences were prepared by H. J. Humpstone and were all taken from a popular current periodical, so as to obtain material of a non-technical character which would be of suitable difficulty and complexity for the ordinary adult. In each pair of sentences, the first is designed to encourage visual imagery, while the second is of a more abstract nature and does not lend itself readily to any type of sensory imagery.

In administering the test, the sentences were read aloud with natural expression, the class having been instructed to reproduce each sentence graphically, immediately following its presentation. The number of syllables in the longest sentence reproduced verbatim was considered the memory span for syllables for each individual. The scores thus obtained were distributed in the usual manner and quintiled.

Discussion: The test here described is an adaptation of the "repeating syllables" test used by Binet and modified by Terman (10) in the Stanford revision of the Binet-Simon scale. It was felt that the sentences used by Terman in the average adult group were not well suited to the college student, and it was also desired to extend the series beyond twenty-eight syllables. Results so far obtained with the Humpstone sentences show a maximum span of forty syllables, a minimum of twenty, with a decided mode at thirty syllables. An analysis of a large number of results has shown no significant difference in the difficulty of the visual and abstract sentences.

The test may be said to measure the integrated memory span. While the factor of associability is probably predominant, the elements in this case are not discrete as in the memory span for digits, and reproduction calls for a higher degree of intellectual organization. Memory is a more important factor than in the span for digits, since the period of retention is somewhat longer, but again it cannot be considered the ability primarily tested. Language ability is certainly involved but the popular character of the sentences employed minimizes its importance when the test is applied to college students. The use of tests of this nature as a measure of proficiency in a foreign language is suggested in this connection. The memory span for syllables must be considered an index of integrability rather than of simple associability.

Memory Span for Ideas.

Description: The paragraph beginning "Tests such as we are now making" from the superior adult series of the Stanford revision was used as the material for the test. The standard directions were given with necessary modification for graphic instead of oral reproduction, as follows: "I am going to read a little selection of about six or eight lines. When I am through I will ask you to write as much of it as you can. It doesn't make any difference whether you remember the exact words or not, but you must listen carefully so that you can write down everything it says." The paragraph was then read at a natural rate, following which adequate time was allowed for reproduction.

The results were scored on the basis of the number of ideas correctly recorded, the paragraph having been analyzed into sixteen discrete ideas. The scores thus obtained were arranged in rank order and the quintile ratings determined.

Discussion: While this test is spoken of by Whipple and others as a measure of logical as contrasted with rote memory, Terman calls attention to the fact that it is rather a test "of ability to comprehend the drift of an abstract passage". It seems more satisfactory, however, to regard the memory span for ideas as a natural sequent to the spans for digits and syllables. It will readily be granted that the college student, who receives most of his mental pabulum through the medium of lectures, can comprehend the drift of such a passage as the one here employed. The test must, therefore, be considered a measure of the subject's ability to associate in consciousness a number of logically related ideas. That this requires a higher level of intellectual organization than the verbatim reproduction of a sentence, as in the memory span for syllables, is hardly open to

question. The test, then, involves not only the element of associability but likewise a high degree of understanding and of intellect. It would therefore be reasonable to expect this test to be more significant when applied to college students than either the memory span for digits or for syllables.

While there may be some disagreement as to what constitutes the unit idea which is to be used as the basis for scoring, the method employed by Terman is too vague for the present purpose, and it is believed that the comparative results obtained by any logical scoring system will be significant.

Description Test.

Description: The Witmer formboard, a modification of that of Seguin, was used as the object to be described. The Witmer board provides recesses for eleven forms, namely the square, rectangle, cross, oval, semicircle, star, equilateral triangle, isosceles triangle, hexagon, circle, and diamond. The following instructions were given: "I have here an object. I am not going to give you a name for it. You can call it a 'thing'—call it 'X'. I want you to pass it around so that each one in the class has an opportunity to examine it." A number of formboards were then passed about the class, and after six minutes had been allowed for examination they were collected and placed on tables in different parts of the laboratory where they could easily be seen. Further instructions were then given. "What is it? In answer to that question I want you to write a description in such a way that anyone would understand and recognize this object. You will be allowed twenty minutes in which to write this paper."

Upon the completion of the twenty-minute period, the written descriptions were collected and redistributed to other members of the class. The number of "points of description" to be used as a basis for scoring of results was then determined in an open discussion. The scores, which were later translated into quintile ratings, were therefore based on an empirical rather than an arbitrary standard.

Discussion: The term "description" as used in this test has reference, not to a literary form, but to the enumeration of the salient characteristics of the object in question. The test is obviously related to the Ausfrage test previously discussed, in that observation is an important factor. In this case, however, memory plays no part, since the object to be described is displayed throughout the twenty-minute period. The test resembles the Aussage test in that no specific questions are asked, the score being based instead on the number of points of description noted. The problem must therefore

be considered one of analysis, and the ability primarily involved may be termed analytic concentration of attention. This ability is contrasted with the distribution or alertness of attention called for in the Taylor number test.

The description test was first used by Binet, who stated that individual psychology can be more readily studied through the examination of complex rather than simple mental processes. The test, in the form of description of pictures, is found in the Binet-Simon scale as well as in the Stanford Revision. When applied to children, the qualitative aspect of the description, whether mere enumeration of points or interpretation, is of more significance than the quantitative score used in this case.

Sentence Completion Test.

Description: Language Scale "K", devised by M. R. Trabue (11), was employed in this test. Owing to its wide familiarity, it is only necessary to remark in this connection that Scale K consists of seven sentences which are arranged in the order of increasing difficulty. Certain words in each sentence have been omitted and the subject is asked to supply the missing words. The procedure standardized by Trabue was adhered to, the following explanation being given before the distribution of the forms.

"This sheet contains some incomplete sentences which form a scale. This scale is to measure how carefully and rapidly you can think, and especially how good you are in language work. You are to write one word on each blank, in each case selecting the word which makes the most sensible statement. You may have just five minutes in which to sign your name at the top of the page and write the words that are missing. The papers will be passed to you with the face downward. Do not turn them over until we are ready. After the signal is given to start, remember that you are to write just *one* word on each blank and that your score depends on the number of perfect sentences you have at the end of five minutes."

The forms were then distributed and the following additional instructions given. "After you have been working five minutes, I shall say, 'The time is up. All stop writing!' You will all please stop at once and lay aside your pens (or pencils). Now if you are all ready, you may turn your papers, sign your names and fill the blanks."

In scoring the results the method recommended by Trabue was followed, a sentence perfectly completed being given two points, one point being allowed where the idea was right but the best word not supplied, and a score of zero received where the completion was

unsatisfactory or omitted. The total number of points for the test was determined and the quintile ratings given.

Discussion: Trabue, in discussing his language scales, does not attempt an analysis of the abilities involved. He calls attention to the fact that the completion test was characterized by Ebbinghaus, who first used the method, as a "real test of intelligence", and that other psychologists have classified it as a test of imagination, memory, association, and various other "faculties". Trabue himself is satisfied with the statement that the "ability to complete these sentences successfully is very closely related to what is usually called language ability".

As has been mentioned by Whipple and others, the ability called into play by the sentence completion test varies greatly with the number and character of the elisions made. If the elisions are few and the nature of the context simple, the problem becomes merely one of controlled association. When the elisions are more numerous the test becomes one of active imagination. An inspection of the seven sentences which form Scale K will show that for the college student the first three sentences and probably the fourth present no imaginative problem, and may be considered comparatively simple tests of controlled association. The remaining sentences, however, are decidedly more difficult, as evidenced by the fact that very few errors were made in the completion of the first four sentences while many were recorded in the fifth, sixth and seventh, and these must be looked upon as tests of imagination. Nevertheless, language ability is of so complex a character, involving as it does various types of sensory imagery, memory, and intellectual organization, that the use of the term imagination in this connection is little more than begging the question.

Although the abilities involved in the sentence completion test are difficult of analysis, the test is of proven significance as an index of "general intelligence", and a study of the nature of the errors made by a subject is often of diagnostic value.

Courtis Arithmetic Test.

Description: The wide acceptance of the Courtis standard tests (12) makes necessary only a brief description here. Series A, Form 3, of the Courtis arithmetic test was used. It consists of a group of eight separate tests in the fundamental processes of arithmetic and their application to problems of varying degrees of difficulty. The first five tests of the series measure efficiency in copying figures, and in simple addition, subtraction, multiplication and

division, respectively. The sixth test requires judgments of the operation to be used in simple one-step problems, and is called by Courtis the speed reasoning test. The seventh, or "fundamentals", test provides abstract examples in the four operations, and serves as a "general measure of the ability to add, subtract, multiply and divide with whole numbers". The eighth test requires judgment of the operations to be used, as well as the actual solution of more difficult two-step problems.

The standard procedure was closely adhered to in administering the tests, one minute being allowed for each of the first six, twelve minutes for the seventh, and six minutes for the last test. After the results had been scored in the usual manner, the scores for each test were treated separately, class distributions being made and quintile ratings assigned. The eight quintile grades for each individual were then averaged and the averages thus obtained were in turn put in rank order and quintiled. This final quintile rating appears in the tabulation of results in a later section of this report.

Discussion: The Courtis arithmetic tests provide a valuable illustration of the efficiency test as contrasted with that of intelligence. Although this is true to a greater degree of the first five tests than of the sixth, seventh and eighth, even these latter must be considered tests of efficiency when applied to college students. It may be assumed that every member of such a group has the educational background and mathematical ability necessary to solve each of the simple problems presented, and the test therefore measures the facility with which the fundamental processes can be employed. It is not the intention here to attempt to analyze the specific abilities involved in arithmetic. It has even been asserted that mathematical ability is itself specific, akin, for example, to musical ability. Certainly, the factor of intellect cannot be disregarded, and in such a test as this, alertness of attention and motor coördination are also important. Since the higher curriculum does not frequently call for exercise in the simpler mathematical operations, it is not surprising to find that the college student often fails to meet the standards of the higher elementary grades. This fact illustrates clearly the distinction between efficiency and competency.

It would be well to note in this connection the service which the Courtis tests have performed in introducing scientific measurements in the field of education. The tests were designed primarily to determine the efficiency of the teacher or of the school system and not to discover individual competency.

Differences and Likenesses.

Description: The tests here referred to are all found in the Stanford Revision of the Binet-Simon scale, and include the "differences" test from Year VII, the "similarities—two things" test from Year VIII, the "similarities—three things" test from Year XII, and the "differences between president and king" from Year XIV group. The Terman method was closely adhered to in giving the tests except that a time element was introduced. One minute was allowed for each part of the seven- and eight-year tests, two minutes for the twelve-year test, and five minutes for the fourteen-year test. The response in each case was written instead of oral. In scoring, one point was given for each correct difference or similarity. It was necessary, however, to quintile the papers largely on the basis of a qualitative judgment of the results, since the tests here described do not present a real problem to the college student.

Discussion: Since the association of ideas with reference to differences and similarities constitutes the essential element of the higher thought processes, these tests are of great significance when applied to children, and were included in this series chiefly for their illustrative value. From a genetic point of view, the recognition of differences is an earlier development than the appreciation of similarities, as evidenced by the Terman standardization which places them at the seventh and eighth years, respectively. However, although similarity in the use of familiar objects should be given at the eight-year mental level, it is not until the twelfth year that the concept thas become usable to the extent of classing the snake, the cow and the sparrow as animals. It is not until the adult level has practically been reached that the ability to appreciate essential differences and likenesses is evident, and this ability may be considered a significant index of intellectual development.

The test in its present form cannot be considered satisfactory for college students, and as Terman suggests it would be advantageous to develop and standardize a new test designed primarily for use in the upper years and at the adult level, and adapted to call into play the ability to give essential differences and likenesses. As a test for adults the one here used can only be said to exercise the associational processes.

Opposites Test.

Description: The difficult opposites found in List V, page 79, of Whipple's Manual of Mental and Physical Tests (13) was used. The directions suggested by Whipple were given, as follows: "Write as soon as I say a word as quickly as you can the word that means

just the opposite. Opposites formed by the prefixes 'un' and 'in' or by the suffix 'less' are not to be given unless the root of the stimulus word is changed." The stimulus words were called at five-second intervals, and the results scored upon the basis of correct opposites determined in open discussion.

Discussion: Tests of controlled association, such as the part-whole, genus-species, and opposites tests, are usually scored on the basis of the time required and the accuracy of the response. In the present case, however, since printed forms were not used, the time element had to be ignored except in so far as the five-second period eliminated all associations requiring a greater length of time. In the scoring of the test a difficulty was encountered in the determination of correct or permissible opposites, and in some cases where no original opposite could be agreed upon the use of two or even three terms was allowed.

The opposites test has been extensively used by Thorndike, Woodworth and Wells, Miss Norsworthy and others. The abilities involved vary considerably with the ease or difficulty of the stimulus words. If the associations called for are too simple the response becomes automatic, while if the stimulus words are very difficult lack of familiarity with the terms is likely to interfere with the validity of the test. It may safely be stated that every word in the list here employed is familiar to college students, and that, with one or two exceptions, the associations required were difficult enough to eliminate automatic responses. It is therefore reasonable to consider the test a measure of the facility and accuracy of controlled association, involving a high degree of language ability.

Definitions Test.

Description: As in the case of the differences and likenesses test, a series of tests from different age levels of the Stanford Revision of the Binet-Simon scale were used. The definitions tests from Year V and from Year VIII, the definition of abstract terms from Year XII, and the differences between abstract terms from the average adult series comprise the present test. The Terman method was employed except that the definition was written and a time element introduced. One minute was allowed for each of the definitions in the first three tests and two minutes for each in the fourth test. In scoring, the same method was followed as in the case of differences and likenesses, and the same criticism as to the accuracy of the quintile ratings applies here.

Discussion: The definitions test differs from those previously discussed in that it tests neither intelligence nor efficiency in mental

processes, but is employed as an index of intellectual development as displayed by the number of words at the disposal of the individual. Since it may fairly be said that formal education consists in adding to the number of usable idea-symbols and increasing their distinction, the vocabulary test provides a simple and quite trustworthy measurement of intellectual status. With formal education so important a factor in each, it is not surprising to find the high degree of correlation noted by Terman between the results of his vocabulary test and intelligence quotients determined by the Stanford Revision.

While the principle involved is the same, the test here employed differs from the usual vocabulary test in that only a limited number of definitions were called for. The purpose was rather to demonstrate the various stages of definition than to actually test the college group. Beginning with definition "by use" at the five-year level, the series shows the development of definition "superior to use" in the eighth year. Both of these types have a definite perceptual basis, and it is not until the twelfth year that the processes of comparison and generalization make possible the definition of abstract terms. In the contrasting of abstract terms, definition is related to the recognition of essential differences, discussed in a previous test.

For the college student even the most difficult test of this series can hardly be said to present a real problem, although in some cases the contrast is not clearly drawn. While such processes as discrimination and classification enter into definition, the test may be considered one of intellectual development as displayed in language ability.

Humpstone Memory Test.

Description: The memory test devised by H. J. Humpstone consists of twenty sentences, each the statement of some rather obscure historic fact connected with the name of some individual or nation. These statements are in the form of the following sentence, "North America was discovered by Columbus in 1492". The series of twenty statements was read aloud to the class three times, care being taken to pronounce the proper names and the dates distinctly. A general discussion, not connected with the experiment, was then entered into and continued for forty minutes. At the expiration of that period, the first part of each sentence was read and the members of the class asked to record in writing the name and date connected with the incident. For example, the experimenter might read "North America was discovered by ——", the remainder of the sentence being supplied by each subject. It should be kept in mind that care was taken in devising the test to select historical incidents

of a trivial and therefore unfamiliar character. Since each of the twenty sentences required the recall of a name and a date, the results were scored on the basis of forty points. The scores were distributed and quintiled in the usual manner.

Discussion: Various types of memory tests have been devised and employed since Ebbinghaus published his pioneer study in this field. Some of these have been open to the criticism that they test associability rather than memory, others that the material is unsatisfactory, as in the case of nonsense syllables, and still others that the time element involved makes them impractical for use in the classroom. The purpose in devising the test here described was to select simple material which at the same time would be unfamiliar, and would offer sufficient points for scoring to provide the necessary differentiation of results. It was further desired to construct a test which might be completed in the two-hour laboratory period and still give sufficient weight to the factor of retentiveness to make the test really one of memory. The Humpstone Test seems to fulfil these requirements satisfactorily. The three readings of the material bring in the element of repetition and give a fair degree of initial memorization. The interval and distraction provided by the forty-minute discussion involve sufficient retention to make the test significant, and the fact that no perfect scores have been made demonstrates that the material chosen is of sufficient difficulty for a college group. The method of right associates employed in the recall needs no comment because of its general acceptance. The natural division of the recalled items into names and dates has shown the latter to be more difficult of retention, as might be anticipated.

It is unnecessary at this point to enter into an analytic study of memory. The subject has been so thoroughly treated in standard text-books and scientific researches as to require no exposition here. It will be sufficient to note that the present test adequately calls into play the three abilities which are chiefly concerned in memory, namely, modifiability, retentivity, and recall.

Witmer Cylinder Test.

Description: The material here employed is an adaptation of the Montessori cylinders, and consists of a circular board containing recesses for eighteen cylindrical insets. These insets are arranged in three series of seven blocks each, the last cylinder of one series being also the first of the next series. In the first series the insets are all of equal diameter and vary only in height, in the second the variation is in diameter, the height being constant, while the cylinders of the third series vary in both height and diameter. The board, which is

approximately twelve inches in diameter, contains a central recess in which all of the blocks may be placed, the subject then being required to replace them as quickly as possible.

Each member of the elementary class in psychology was tested individually by either Professor Witmer, Professor Twitmyer, or Dr. Humpstone, this being the only one of the series which was not given as a group test. The student was required to stand before the table upon which the cylinder board was placed with all of the insets in their proper recesses. His attention was called to the fact that the tops of the different blocks were flush with the top of the board. The insets were then removed by the experimenter and placed in the central receptacle, care being taken to mix the blocks well and at the same time to leave the larger cylinders on top. The subject was then instructed to return the blocks to their original positions as rapidly as possible, and the time required was recorded in seconds. Upon the completion of the first trial the cylinders were again removed and the time for the second replacement determined.

The results for each of the two trials were treated separately and quintile ratings obtained. In accordance with the method standardized by Paschal (14), a final rating was given by quintiling the results for the shortest trial. The rating for the first, second, and shortest trials all appear in the tabulation of results.

Discussion: While the diagnostic value of the mechanical test has long been recognized, the cylinder test is the only one of this type which has been included in the present series. The test differs from those which have previously been described in not requiring any appreciable degree of language ability, and hence can not be considered in any sense an index of intellectual level. If intelligence be defined as the ability to solve what for the individual is a new problem, the test is primarily one of intelligence. This, however, is by no means the only ability involved. On the motor side may be observed the rate of discharge of energy, coördination, complexity of response, and in some cases endurance. The performance likewise displays some degree of analytic and distributed attention, observation, understanding, and trainability when more than one trial is given. While these are not the only abilities involved, they may usually be rated with some accuracy on the basis of the cylinder performance.

As Paschal has pointed out, the test has both a qualitative and a quantitative aspect. In the present treatment of results, however, only the latter has been considered, since the performance has been rated solely on the basis of the number of seconds required for the successful replacement of the insets. The qualitative aspect of the

performance was an important factor in determining the competency rating which will be discussed in the following section. In general, the quality of the performance must be considered of more diagnostic significance than the bare time element, although it is evident that a very rapid replacement can not be made unless the performance is qualitatively good, nor is it likely that excessive time will be required if a satisfactory method is used.

While the quintile ratings for the first and second trials have been included in the tabulation of results, it is probable that the rating for the shortest of the two trials gives the safest index of cylinder proficiency. In his standardization of the test Paschal adopted the shortest of three trials as his criterion, and the results here obtained are therefore not directly comparable with those upon which the standardization was based. Even though the shortest trial gives the most reliable basis for a single rating, the comparison of scores made on the first and second trials is important as an index of trainability, and these have therefore been included in the table of results.

Composite Test Rating.

In the treatment of results it will be of interest to compare the records made in the various tests described above with the score of the Thurstone test, the rating on academic standing and that on estimated competency. It seems advisable to obtain a composite rating on the basis of the results for the series of tests in order to facilitate this comparison. Unquestionably, the tests are not all of equal value, and some method of weighting should be employed. Here, however, an almost unsolvable problem is encountered, for any system of weighting the various tests which might be adopted would necessarily be arbitrary and based on an *a priori* judgment. Moreover, the significance of the tests varies with the individual case, and no one method of weighting would be really satisfactory for the whole group.

With these difficulties in mind, it has been decided to obtain a composite rating by taking a simple average of the quintile scores on the thirteen tests of the series for each individual. Such an average has the advantage of not being colored by personal opinions of the value of the different tests, and is probably as significant an index as could be devised by any complicated system of weighting. This average includes only the rating for the shortest trial with the Witmer cylinders.

The Competency Rating.

One purpose of this investigation was to determine what reliance may be placed on the "snap judgment" of a trained observer. Is it possible to rate the college student with any degree of accuracy on the basis of an interview covering no more than five minutes? Can the experienced psychologist estimate the ability of an individual by noting his appearance and carriage, and by obtaining his reaction to a few simple questions and observing his performance with a mechanical test? It was with a view to answering such questions as these that each member of the first-year class in psychology was personally interviewed by either Professor Witmer, Professor Twitmyer or Dr. Humpstone, and given a competency rating on the basis of five minutes' observation. Each student was required to replace the insets of the Witmer cylinder test twice, as described in the preceding section. The qualitative aspect of this performance had considerable weight in determining the competency rating, and it should be understood that while coördination, attention, understanding, trainability and intelligence are all reflected in the time scores of the two cylinder trials, the latter do not necessarily correlate with a rating based on the quality of the performance. As has been previously noted, the cylinder test is the only one considered in this study which was given individually.

The rating, however, was not based solely on the performance with the cylinders. As the student presented himself to the examiner, he was asked to write his name upon a record card, and the character of his writing as well as the degree of composure displayed were observed. A few leading questions were then asked regarding preparatory school, purpose in coming to the University, intended vocation, outside activities, and the like. No attempt was made to ask the same questions of each individual, but rather to carry on a short conversation which varied naturally with the replies given. The subject was then given the cylinder test, following the procedure previously outlined, and after answering one or two questions as to his work in psychology was dismissed. As a rule the whole interview consumed no more than five minutes.

While all three of the examiners had come into some contact with members of the class through lecture work, no one of them knew the students personally or had had occasion to be familiar with the type of work done by any individual. The rating was therefore based entirely upon an observation of the student's behavior as displayed in his general bearing and address, his answers to the questions, and his performance with the cylinders. In this respect, the competency rating here employed differs from the rating on estimated intelli-

gence which has frequently been used in connection with investigations of this character. Such a rating has usually been given by an instructor familiar with the student and with his work in the classroom, or by averaging the estimates made by a number of instructors so qualified. The competency rating is therefore not directly comparable with the ratings on estimated intelligence referred to in a preceding section.

In giving these ratings, the five-point scale was used in a somewhat modified form. Each of the five points of the scale was subdivided into five lesser grades, thus giving a maximum rating of 5.5, a minimum of 1.1, and a mediocre grade of 3.3. When each student had been rated on this scale, the three examiners in conference arranged the members of the class in rank order on the basis of estimated competency. Since it is felt that individual differences in the standards of the three examiners somewhat reduces the significance of the actual rating assigned, the rank order has been employed in determining a quintile rating on estimated competency, which appears in the tabulation of results. This treatment has the added advantage of making the rating directly comparable with the quintile scores of the various mental tests.

It will be well to note at this point that there is no objective standard by which to measure the accuracy of the competency ratings. In estimating the ability of the student, the attempt was not made to predict the degree of his success in the study of psychology, nor is the rating a prognosis of his relative academic standing as determined by the grades received in all college courses. Neither can the accuracy of the judgment be measured by his performance in any one or in any group of psychological tests. The term "competency rating", implying the algebraic sum of the individual's specific abilities and disabilities as demonstrated by his success as a member of human society, best interprets the character of the rating under discussion. In this connection it may be stated that no ratings lower than 2.3 were given, or, in other words, no students were found so deficient in general competency as to fall below the "doubtful" group. In view of the fact that the members of the class had undergone a strenuous process of selection in fulfilling the entrance requirements and surviving at least two years of college work, it is not surprising to find a complete absence of "1" ratings. This fact does not appear in the tabulation of results where the ratings have been quintiled on the basis of rank order, and only the quintile grade shown.

Although, as has been pointed out, the competency rating can not be checked by comparison with mental test scores or academic record, it will nevertheless be profitable to determine in the later

treatment of results whether the rating shows any significant degree of correlation with competency as displayed in the tests and college grades.

Academic Rating.

Popular tradition has it that the youth whose scholastic attainments make him valedictorian of his college class is destined for future mediocrity, while the typical campus lounger whose academic life is cut short by a heartless faculty is sure to make his mark in the world of success. Nevertheless it will hardly be denied that proficiency in the classroom is to some degree indicative of individual competency, and it will therefore be desirable to know something of the relative academic standing of the fifty students under consideration.

While it might be contended that preparatory school records would be significant in determining a rating on scholastic merit, the great variation in standards and the incomparability of the various grading systems employed make it advisable to reject this suggestion without further deliberation. Moreover, since grades for at least two years of college work are available for each member of the group, it seems unnecessary to base the academic rating on any work other than that done at the University of Pennsylvania.

As has previously been stated, a five-point system of grading is employed in the School of Arts and Science. This scheme provides three symbols for work of passing grade, while two are reserved for that of an unsatisfactory character. To be more specific, the letters "D", "G", "P", "N", and "F" are assigned, signifying Distinguished, Good, Passed, Not Passed (conditioned), and Failure, respectively. A student receiving a grade of "N" in a course may relieve himself of the condition by passing a re-examination, while the grade "F" necessitates the repetition of the course. As applied to the courses in psychology, an "N" in Psychology 1 permits the student to continue his work in Psychology 2, but this permission is not given when "F" is received in the first course. It will be noted, therefore, that no member of the present group received a grade of "F" in Psychology 1, since each of the fifty students completed both courses in the academic year 1919–20.

While it must be understood that the letter system of grading is intended to obviate the pseudo-accuracy of the percentile grade, and that it is not possible to assign percentile equivalents for the symbols used, the necessity for obtaining some kind of composite rating as an index of academic standing is evident. For example, a given student may have received a grade of "D" in five courses,

"G" in eight, "P" in four, and "N" in two. Moreover, each course may have required from one to nine hours of class attendance per week, with a value of from one to four units of credit, a unit being the equivalent of one hour of lecture work or two hours of laboratory work per week for the academic year. Hence it is clear that the grades must be considered in terms of units of credit rather than by courses if a significant rating is to be obtained, and also that some numerical translation of the letter grades must be devised.

Since the percentile scale is not recognized in the University marking system, any numerical equivalents which might be adopted would necessarily be arbitrary. Roughly, it may be said that "D" represents a range from 90 to 100 per cent, "G" from 80 to 90, and "P" from 70 to 80. There seems to be no justification, however, for selecting 70 per cent as the marking mark, nor would it be more accurate to place it at 60 per cent. A satisfactory evaluation of the "N" and "F" is even more confusing. While the passing grades might be valued at 95, 85, and 70, respectively, it would be difficult to decide whether the "F", which ranges from zero to 50 per cent should be rated as 25 or 45. By far the simplest solution to the problem, and what seems to be the most logical, is to adopt here the five-point scale generally employed in this study. It is quite as reasonable to represent the five-letter grades by the numbers 5, 4, 3, 2, and 1, as by any other numerical values which might be suggested, and this method has the advantage of permitting a direct comparison between the composite rating for college grades and that for mental tests. It has been determined, moreover, that the rank order remains approximately the same whether this system is used or the values 95, 85, 70, 55, and 45 be given to the letter grades.

The academic rating has therefore been determined by multiplying the number of units assigned each letter grade by the appropriate digit, and dividing the sum of these products by the total number of units graded. The student who had received no grades lower than "D" would have a rating of 5.0, while a record with an equal number of "G" and "D" units would average 4.5. Since it would be almost impossible for a student to remain in college who had not averaged the passing grade, it is not surprising to find that only one of the fifty has an average below 3.0, his rating being 2.9.

Perhaps even a rating of this kind implies an accuracy of measurement which cannot be justified. If every "D" assigned as a final grade stands for the same level of excellence, and if the same amount of work is required for a passing grade in every course, then the validity of the average rating cannot be questioned. If, however, one department of the college is found to be giving the highest grade

to 25 per cent of its students, while a second allows only 5 per cent of "D"s, then the impossibility of comparing grades assigned by different departments is evident. Moreover, it has been demonstrated that different instructors in the same department vary greatly in the grades which they assign to a given piece of work, and that this variation is no greater than that which will be shown by one instructor marking the same work at different times. It is indeed questionable whether any reliance should be placed in a comparison of college grades in an institution where the majority of the courses are elective, and where there is no general supervision of grading.

The grading problem is by no means a new one, and has a considerable literature of its own. Finkelstein (15), for example, has published an interesting study of conditions at Cornell University in which he demonstrates the need for supervision of the grades assigned by different departments by showing that some instructors are typically low markers. He makes a plea for the adoption of a five-division system of grading with the provision that the grades given by any instructor shall not deviate in the long run from a distribution agreed upon. While the intention of the present study is not to preach the necessity of some such general supervision of grading at the University of Pennsylvania, the existing absence of uniformity demands comment. Under the present curriculum, a student in the School of Arts and Science is required to complete a specified number of units of work in each "group" of subjects. In most cases he is free to elect which courses he will pursue in a given group. For example, six units of credit is required in the Biological Science Group which is composed of courses in botany, zoölogy and psychology, but the decision as to whether all six units be taken in one subject or be distributed between two is left entirely to the student, as well as the choice of the subject or subjects to be elected. Until recently the elementary work in one of the three subjects has been so much less difficult than that in the other two, that the situation has been fully recognized by the undergraduate, with a consequent influx to the easier course. While this condition has been remedied in the case cited, it doubtless still exists in other groups, and the present plea is made rather with the purpose of calling attention to the lack of general supervision of grade distributions than as a criticism of any particular instance of non-conformity. Although the necessity of some general supervision of all grades assigned in the college cannot be overlooked, the more pressing need of uniformity within the various groups must be emphasized.

From the foregoing discussion it is evident that grades assigned by various instructors in different departments of the University are not really comparable, and it is with this understanding that the

academic record will be included in the present investigation. Even though the data cannot be considered scientifically accurate, however, it must be admitted that the student's college grades do give some indication of his scholastic ability. The grades alone determine whether he is to receive academic honors or be dropped by the Executive Committee for general deficiency, as well playing an important part in election to Phi Beta Kappa and in placement after graduation.

In the tabulation of results, the final grades for the two courses in psychology will be noted in addition to the average rating for all college grades including those in psychology. The latter are given separately since it is felt that the unusual opportunity for personal contact between instructor and student in the elementary courses in this department makes these grades somewhat more significant than is generally the case.

In conclusion, it seems almost unnecessary to point to the fact that similar grades may not mean the same thing when assigned to different students even in the same course. Although the attempt is made to control the amount of work done by fixing the maximum as well as the minimum number of units which may be taken by a student in a semester, some carry so full a roster as to seriously interfere with the display of actual ability, while others who are not experiencing great success with a comparatively light schedule may be handicapped by outside work which they are pursuing as a means of livelihood. Since the evaluation of these distributing factors is well nigh impossible, they must be ignored in the present treatment of college grades.

Tabulation of Results.

While it was intended to make a statistical study of the various scores and ratings which form a basis for this investigation, the primary purpose was to study the individual record rather than the mass results. It has therefore been deemed advisable to present a complete tabulation of the ratings for each member of the group, and thereby facilitate the scrutiny of the individual case. In the following table will be found (1) the number used to designate each student in the group, (2) his class, whether sophomore, junior, or senior, (3) the quintile rating for the Thurstone test, (4) the quintile rating for each of the thirteen mental tests with the addition of the ratings for the first and second trials with the cylinders, (5) the composite test rating obtained by averaging the ratings for the thirteen separate tests—this average does not include the Thurstone test and only the shortest trial with the cylinders is included, (6) the quintile

Tabulation of Results.

Student No.	Class	Thurstone Quintile	Ausfrage	Taylor No.	Memory Span			Description	Trabue Scale	Courtis "A"	Diff. and Like.	Opposites	Definitions	Memory	Cylinder	Cylinder 1st	Cylinder 2d	Average Test Rating	Competency Rating	Academic Rating	Psychology 1	Psychology 2
					Digits	Syllables	Ideas															
1	So.	3	4	2	1	3	3	3	2	2	4	4	4	4	2	1	2	2.9	3	4.5	G	D
2	So.	5	4	4	5	4	5	4	4	5	3	3	3	3	4	4	4	3.9	5	4.7	D	D
3	So.	2	1	3	3	3	2	2	5	2	3	3	2	1	4	5	2	2.6	4	3.4	N	N
4	So.	5	2	3	2	3	5	2	5	5	3	4	5	–	1	2	1	3.3	3	3.6	G	P
5	So.	3	5	4	1	3	1	3	4	5	3	3	1	1	5	2	5	3.0	2	3.9	P	P
6	Jr.	5	5	3	5	3	2	–	5	2	3	4	2	4	1	1	1	3.3	1	3.8	P	P
7	So.	4	2	3	2	3	5	3	1	4	3	3	4	2	5	5	5	3.1	5	3.2	N	F
8	So.	4	4	1	3	4	3	2	5	2	3	4	1	3	2	2	2	2.8	3	3.6	P	G
9	So.	4	2	3	3	3	3	4	1	2	3	3	3	4	5	5	5	3.0	5	4.1	G	G
10	So.	1	2	4	2	3	4	3	1	2	1	2	2	4	3	4	4	2.5	4	3.5	P	P
11	Jr.	1	3	3	5	3	2	4	1	5	3	3	5	4	4	1	4	3.5	1	4.2	P	G
12	So.	4	3	4	1	3	3	3	5	2	3	3	2	4	3	2	3	3.0	3	4.1	P	G
13	So.	4	3	5	5	3	4	3	5	4	2	–	2	1	5	5	5	3.5	5	4.5	G	D
14	Sr.	5	5	2	5	4	3	3	5	4	3	3	3	5	3	1	3	3.7	4	4.2	G	G
15	Sr.	5	5	4	3	4	5	3	–	2	3	3	2	3	5	4	5	3.5	5	3.5	P	P
16	So.	4	5	2	2	2	3	3	1	4	3	3	2	4	3	3	3	2.8	2	3.7	P	P
17	So.	1	1	2	4	3	2	5	1	1	3	1	4	2	1	3	1	2.3	1	3.7	P	G
18	So.	5	3	3	3	3	5	3	3	5	3	3	2	4	3	3	3	3.3	3	4.5	P	D
19	So.	5	3	3	5	4	5	3	5	2	3	3	5	4	1	3	1	3.5	4	3.9	P	P
20	So.	5	3	4	3	4	3	3	5	1	3	3	5	1	1	2	1	3.0	4	3.4	P	P
21	So.	4	4	4	3	3	4	3	4	5	3	3	2	3	4	2	4	3.5	4	4.2	P	D
22	So.	2	–	–	3	3	3	1	3	3	3	4	4	5	1	1	1	3.0	1	3.5	P	P
23	So.	3	2	4	2	3	3	4	1	5	3	3	3	5	3	4	4	3.2	4	3.7	P	N
24	So.	3	5	5	5	3	4	3	3	3	3	3	2	4	5	2	5	3.7	4	3.1	P	P
25	Jr.	3	5	2	2	2	3	3	5	4	–	3	2	4	3	4	3	3.2	3	3.3	N	P
26	Jr.	1	4	3	1	4	4	3	3	3	3	2	1	2	5	3	5	2.9	3	4.0	P	P
27	So.	5	5	4	5	4	5	3	5	3	3	4	2	4	3	4	3	3.8	3	4.4	D	D
28	So.	3	5	5	3	3	–	3	2	5	4	5	4	4	2	2	2	3.8	3	4.7	D	D
29	Jr.	4	5	3	3	3	5	3	2	5	3	3	2	3	4	5	4	3.4	5	4.0	D	G
30	Sr.	4	5	4	2	3	4	–	3	3	3	4	1	5	5	3	5	3.5	3	3.7	P	N
31	Jr.	4	–	5	2	4	3	1	4	2	2	3	–	3	4	3	4	3.0	4	4.0	G	P
32	So.	5	3	2	2	4	3	3	1	3	1	3	2	5	1	3	2	2.5	2	4.6	D	G
33	So.	1	3	1	2	1	3	3	2	4	2	2	1	5	3	1	3	2.5	1	3.1	P	N
34	Sr.	5	2	2	5	3	3	–	3	4	3	3	3	5	3	3	4	3.3	4	4.3	D	G
35	Jr.	5	3	5	5	3	3	3	5	4	3	3	4	5	5	5	5	3.9	5	3.7	G	G
36	Jr.	4	4	3	5	4	–	–	3	4	–	–	2	–	2	3	2	3.5	4	3.0	G	P
37	Jr.	2	4	4	4	3	3	2	3	4	3	–	2	2	5	4	5	3.3	3	3.4	P	G
38	So.	2	–	2	4	3	2	2	1	3	3	3	4	3	4	4	4	2.8	2	3.0	P	P
39	So.	1	4	4	3	3	3	1	2	1	2	3	3	2	4	5	4	2.7	2	3.1	P	P
40	So.	4	1	3	2	4	4	3	3	4	2	3	4	5	1	1	1	3.0	1	3.9	G	P
41	So.	3	5	1	2	2	3	3	2	5	3	4	4	2	5	5	1	3.2	5	3.5	G	G
42	So.	5	5	3	2	3	4	5	5	2	3	2	4	4	5	5	5	3.6	5	4.5	G	D
43	Jr.	3	5	2	2	3	3	3	5	2	3	3	5	3	1	1	1	3.1	1	3.2	P	P
44	So.	3	5	4	2	4	4	3	4	3	3	–	4	4	3	5	3	3.6	4	4.2	G	G
45	So.	3	1	5	2	2	3	2	1	3	3	1	4	3	1	2	1	2.4	2	4.1	P	P
46	Sr.	2	4	4	2	3	2	3	3	5	4	4	4	–	5	5	5	3.6	5	2.9	P	N
47	So.	5	4	2	3	3	4	3	4	5	4	2	–	–	1	1	1	3.0	2	3.5	N	F
48	Jr.	5	5	–	5	4	3	–	1	1	5	3	5	4	5	5	5	3.7	5	3.4	G	P
49	So.	2	4	2	2	5	1	1	2	2	2	3	3	4	2	4	1	2.5	2	3.7	P	N
50	Jr.	5	2	1	3	4	3	1	1	1	3	2	2	5	3	1	3	2.4	3	3.9	G	P

rating based on estimated competency, (7) the final grades in Psychology 1 and Psychology 2, (8) the academic rating obtained by averaging college grades as previously described.

In studying the tabulation of results it must be borne in mind that in every case the quintile rating was obtained from the distribution of the results of the class of approximately two hundred students, and not merely on the basis of the fifty here included. This explains the fact that the ratings are not equally divided among the five quintiles.

Discussion of Results.

In considering the data tabulated on the preceding page, it will first be of interest to determine whether any significant correlations exist between the various ratings given for the group as a whole, and then to study the results for the individual student. It will be valuable to ascertain, for example, whether the rating for the Thurstone test correlates with the average score for the series of more specialized mental tests. Since general intelligence may be looked upon as an average of the specific abilities of the individual, a high correlation might well be expected between these two ratings. Each of these, in turn, must be compared with the rating on estimated competency, and it will likewise be profitable to observe whether any one of these three ratings may be considered an index of proficiency in college work.

With this purpose in view a series of intercorrelations has been calculated between the ratings assigned for the four general divisions of the results. In each case the coefficient of correlation was obtained by the Pearson method. The data employed consists of the quintile grade on the Thurstone test, the average rating for the thirteen mental tests, the quintile rating on estimated competency, and the average rating for college grades.

Correlations.

Competency rating with mental tests.....	$r = +0.49$
Thurstone test with mental tests.....	$r = +0.40$
Thurstone test with college grades.....	$r = +0.39$
Thurstone test with competency rating....	$r = +0.36$
College grades with mental tests.....	$r = +0.21$
College grades with competency rating....	$r = +0.10$

A mere inspection of the coefficients listed above will show that while all of the correlations are positive, not one can be considered

significant. In general, it may be stated that coefficients between +0.30 and +0.75 show that the same factors are operative in the two series to some degree, but the correlation can hardly be regarded as significant unless a coefficient greater than +0.75 is found. An immediate conclusion can therefore be drawn either to the effect that the values employed are not to be relied upon, or that the performances rated in the four cases did not involve the same factors or abilities. Nevertheless, it will be of interest to scrutinize the coefficients obtained more closely, and to attempt to interpret them.

The highest correlation of the series is found to exist between the rating for estimated competency and that for mental tests. This is not surprising since the competency rating was given largely on the basis of the performance displayed in the solution of one of these tests. In view of the fact that the cylinder test calls into play so many of the abilities which enter into other tests of the series, it is rather surprising that the correlation did not prove greater. This can probably be accounted for by the fact that the cylinder test does not involve language ability, which is an important factor in practically all of the other tests.

Next in order is found the correlation between the Thurstone test and the mental test rating. As has been pointed out, both of these ratings may in a sense be considered indices of general intelligence, and since many tests in the series involve intellectual processes similar to those called for in the Thurstone examination, the low correlation displayed here is again unexpected. However, the weight given to the time element in the latter test is so great, and the range of abilities involved so much more restricted than in the Pennsylvania series, that it is not difficult to account for the seeming inconsistency of the results.

The very low correlations obtained between the academic rating on the one hand and the mental test and competency ratings on the other, provide food for serious reflection. The question which must naturally arise is whether academic proficiency, as it is evaluated in our colleges today, is really an index of the competency of the student. Perhaps it will be well to notice whether the low correlation shown here is typical of other similar investigations. In the report by Caldwell (7) previously referred to, appears a summary of the results obtained by other experimenters showing correlations obtained between various series of mental tests and college grades. In this connection, it is unnecessary to note in detail the character of the tests used by each of the investigators, and merely a statement of the correlations obtained, as cited by Caldwell, is shown below.

Correlation of Test Results with College Grades.

Wissler	0.09
Calfee	0.23
Rowland and Lowden	0.37
Waugh	0.41
Kitson	0.44
King and McCrory	0.39
Caldwell	0.44

While the correlations above are in most cases greater than that obtained in the present investigation, namely 0.21, it will be noted that not in a single instance was a significant coefficient shown. Rogers (8) does not even attempt to calculate a coefficient of correlation between test results and college grades, but states that "to predict an individual's probable status in academic work from his performance in the tests would obviously be rash". As has previously been stated, a comparison of the competency rating with ratings on estimated intelligence cited in other investigations is hardly possible, since in this case the estimate was made by an individual unfamiliar with the students rated. It is well to note, however, that even where intelligence was graded by instructors well acquainted with their students, correlations with college grades have not exceeded 0.60.

From the facts given above it is possible to arrive at three conclusions. In the first place, college grades may not actually reflect the mentality of the student, or secondly, the tests employed are inadequate or misleading, or finally, the factors which enter into the assignment of college grades are not the same as those which are measured in psychological tests. Probably all three of these conclusions are in some degree justified.

Voice has been given recently to much criticism of the present university curricula on the grounds of impracticality and because of the continuation of secondary school pedagogical methods in institutions of higher learning. On the other hand, a large proportion of the instruction in our colleges today is given by means of lectures. The grade assigned at the end of the course is often determined chiefly by the student's ability to give back on an examination paper certain information which has been fed to him in lectures during the term. Frequently, little intelligence is called for and the student is rated either on the excellence of his memory or on the degree of industry with which he compensates for a deficiency in that ability. When to this criticism of university instruction is added the unreliability of the grades themselves, as discussed more fully in an earlier section, it is evident that the low correlation between college grades

and test results may be in part due to shortcomings of the educational system both as regards methods of instruction and grading.

In scrutinizing psychological tests as a whole or the series employed here in particular, certain criticisms must be made. Perhaps the exaggeration of the importance of the time element is the most serious fault with the majority of mental tests. Intellectual dexterity is generally measured rather than organization and usability of knowledge. The difficulty is increased in this case by the homogeneity of the group tested. Many of the tests would be significant when applied to individuals less carefully selected and at a lower level of mental development. In most cases the problem presented is too easy to tax the college student, and the speed of reaction is the only ability measured. Another criticism which may be made of tests in general, is that they do not measure with sufficient accuracy the abilities which they are designed to gauge. In other words, a subject does not always give the same score on the same or equivalent tests due to variations in attention, interest, physical condition, etc. Mental testing will not be scientifically accurate until the technique has been so refined as to greatly reduce the probable error of the score, or until a higher reliability coefficient can be obtained. The low correlation between college grades and mental tests may, then, be due to shortcomings of the latter as well as to inaccuracies of the former.

It seems reasonable, however, to believe that this lack of correspondence can be attributed largely to the fact that college work involves other factors than those measured by any series of psychological tests which has yet been devised. In addition to the mental abilities which go to make up the competency of the individual, the factor of motivation plays a most important rôle in academic success. It is possible to conceive of two students of approximately equal competency, one of whom is inspired by the desire to excel in intellectual pursuits, while the other is in college for the purpose of enjoying social or athletic advantages. The intense interest and industry of the first is likely to result in a higher academic rating than would be predicted from his performance in series of mental tests, while just the opposite is true in the case of the second student. While it is fair to believe, therefore, that psychological tests can be employed to select those students who have the ability to succeed in college, they will not form an adequate basis upon which to predict academic success until some means has been devised of measuring motive in quantitative terms. The final solution of the problem will be reached when more accurate methods of assigning college grades have been adopted, and those grades depend more on the

higher thought processes and less on memory, and when, on the other hand, psychological tests have been made more difficult, place less stress on the time element, and include some index of motivation.

Althought it must be admitted that the formulation of a series of mental tests which will accurately predict success in college work is desirable, no great benefit would accrue thereby either to the science of psychology or to the field of education. The psychologist is not so much interested in the abilities which determine college grades, as in evaluating the particular mental assets and liabilities which characterize the individual. While the general intelligence rating, which represents the summation of the scores in a number of tests, is doubtless of some significance, the analysis of such a rating so as to show the peculiar abilities and disabilities of the individual is of much greater importance from the point of view of psychology. An inspection of the results shown in the preceding tabulation reveals the fact that although two students may have the same average test rating, the scores obtained in the different tests are not really reflected in this average. Of two individuals who had an average rating of 3.3 in the thirteen tests and who received the same quintile rating on the Thurstone test, one was placed in the third quintile in nine of the tests, the other in only three. Obviously the first student showed consistent mediocre ability, while the second displayed considerable variation in the different tests, having four ratings of "5", three of "2", and one "1". There is no doubt that the latter student provides the more interesting material for psychological study and for vocational guidance.

Since it is believed that the present investigation is significant rather in the analysis of individual competency than in the correlation of group results, it will be the purpose in the following section to scrutinize the record of each member of the group and to determine whether any conclusions of value in diagnosis or guidance can be reached. In considering the academic rating it is well to note that ratings higher than 4.0 are very good, while those below 3.5 are poor. The median academic rating for the group is 3.7. Composite test ratings above 2.9 and below 3.5 are considered mediocre, with the median rating at 3.2.

Analysis of Individual Records.

No. 1.

This student shows a consistently mediocre record until his college grades are observed, when he is found to have one of the highest academic ratings of the group. Placed in the middle quintile in the Thurstone test as well as in estimated competency, his average

test rating is well below the median. As for the separate tests, he has received the highest rating in none, and the lowest only in the memory span for digits. In general, the higher scores are exhibited in those tests which involve language ability and memory, and the lower where these factors are not prominent, namely, in the Taylor number test, the Courtis test, and the cylinders. In view of the high grades in psychology and the high academic rating it seems probable that this student has some strong motive, such as ambition, and supplements a mediocre intellect with an unusual amount of industry.

No. 2.

This record shows the most consistently high rating to be found in the group. The academic record is the highest, and this is borne out by "distinguished" grades in both courses in psychology. The ratings for estimated competency and for the Thurstone test are both in the fifth quintile, while the average test rating is equaled by only one other student in the group. In considering the results of the particular tests, it will be observed that this student has not fallen below the middle quintile, but has reached the highest in only three tests. He shows the poorer scores in those tests which stress language ability and memory, and the higher ratings where intelligence, imagination and attention are involved. The general level of performance is so high as to make any specific recommendation or prognosis unsafe.

No. 3.

The chief point of interest in this case is the lack of correspondence between the competency rating and the remainder of the data at hand. This student shows an academic rating which places him in the poorest fifth of the group, with conditions for both courses in psychology. Although in the second quintile in the Thurstone test, his average test rating is one of the lowest recorded. He rates above the middle quintile only in the sentence completion and cylinder tests. This may indicate good intelligence not directed toward college work, but the conclusion that the competency rating is too high in this case seems justified.

No. 4.

The indication here is of a student of somewhat more than average general intelligence whose record is largely influenced by interest in the task at hand. With an academic rating slightly above the average and a "G" and "P" in psychology, his score in the Thurstone test puts him in the highest quintile. The composite test rating is slightly above the average, and shows a preponderance of

"5's" as well as a number of "2's" and a "1". High ratings in the memory span for ideas, the sentence completion and the definitions tests, as contrasted with a very poor cylinder performance, indicate intellectual ability rather than intelligence.

No. 5.

This record shows a student somewhat below the average in competency, with an academic rating slightly better than would be expected from the test results. Passing grades in both courses in psychology, estimated competency in the second quintile, and the median rating for the Thurstone test all indicate mediocre ability. This is borne out by an average test rating below the mean for the group. Performances in the memory spans for digits and ideas, and in the definitions and memory tests were rated in the lowest quintile. High scores were obtained in the Ausfrage and Courtis tests and in the second trial with the cylinders. Although the test results show great variation, there seems to be no definite tendency displayed.

No. 6.

This individual probably possesses mediocre ability, although receiving a very low competency rating and a very high score on the Thurstone test. A fair academic rating with passing work in psychology, and a test rating slightly below the average seem to indicate that neither the Thurstone test nor the competency rating gives a true picture of the student. High scores in the Ausfrage, digit span, Trabue, opposites and memory tests, with very poor cylinder performances, would suggest fair intellect coupled with rather deficient intelligence.

No. 7.

The record here indicates relatively low competency with a high degree of native intelligence. A very poor academic rating is substantiated by a condition and a failure in the two courses in psychology, and a low average test rating. An exceptionally good performance with the cylinders and a high rating on the Thurstone test and the idea span, with lower scores on the test requiring language ability and memory, lead to the conclusion that this man is misplaced in college, but would probably succeed in a pursuit which does not stress intellectual development.

No. 8.

There are no outstanding features in the record of this student. The academic rating and the Thurstone score are both slightly above the average, while the test rating is somewhat below. The tests which emphasize the intellectual side usually show good scores,

while those which do not depend on language ability, such as the Taylor number, the Courtis, and the cylinder tests, are placed in the lower quintiles. On the whole, the record is mediocre.

No. 9.

In this instance, a high academic rating, good work in psychology, a high competency rating and a good score on the Thurstone test fail to correlate with a rather low test rating. Median scores on seven of the tests, with only one result in the highest and one in the lowest quintile, indicate a rather consistent mediocrity. A high rating in the memory test and an excellent cylinder performance suggest that good memory and intelligence are responsible for the high academic standing.

No. 10.

A competency rating of "4" indicates that this man was not doing his best on the mental tests. Mediocre college work and a low rating on the Thurstone test suggest that the competency rating is too high. The test scores are generally low where language ability is involved, and are above the average for the Taylor, idea span, memory and cylinder tests. As in Case 7, it seems likely that this individual is not profiting by his college course and would be more successful in some other line of activity.

No. 11.

The record of this student is quite inconsistent. Placed in the lowest quintile in the Thurstone test and competency rating, his test and academic ratings are well above the average. The low score in the first cylinder trial indicates a lack of intelligence, while the marked improvement on the second trial indicates good trainability. The low rating on the Trabue test and idea span, contrasted with high ratings for the Courtis and Humpstone memory tests, suggest an efficient and retentive mind rather than a quick and imaginative one. That this man is a slow thinker is demonstrated by his score on the Thurstone test. The fact that he retains and digests the information which he acquires is evidenced by his high academic record.

No. 12.

This student displays a record consistently near the average for the group. The Thurstone and academic ratings are somewhat better than the mean, the competency rating is in the third quintile, and the test rating slightly below the average. Of the separate tests, seven are rated in the third quintile, a poor score on digit span and a very high rating on the Trabue test being the only significant scores.

On the whole, the competency rating seems to express the ability of the student adequately.

No. 13.

The record in this case is consistently high. Very good grades in the two courses in psychology substantiate an academic rating which is exceeded by only three members of the group. A competency rating of "5" and a Thurstone rating of "4" correlate with a high test rating. The only rating in the lowest quintile is that on the memory test and when this is contrasted with an exceptionally good performance with the cylinders, it seems reasonable to conclude that this student depends more on intelligence than on memory in his college work. Almost without exception ratings in the upper quintiles are displayed for the tests which do not stress language ability, while lower ratings are found where this factor is of great importance.

No. 14.

This record presents an interesting contrast with that of student No. 13 in that the intellectual rather than the intelligence factors are here stressed. While not quite so good from the academic viewpoint, this record shows a slightly higher rating for the Thurstone and other mental tests than does the preceding case. Ratings of "5" on the Ausfrage, digit span, Trabue, and memory tests indicate associability, language ability and retentiveness, while a rating in the lowest quintile for the first cylinder trial implies comparatively poor intelligence. A much better record on the second trial with the cylinders shows trainability, which, coupled with a high memory span and good memory, pictures a student of more than average intellect.

No. 15.

The indication here is of a man of high general intelligence who does not care to apply himself to college work. On the one hand his academic rating is mediocre and he has obtained merely passing grades in psychology, while contrasted with this are Thurstone and competency ratings in the highest quintile, and a combined test rating well above the average. The low rating on the Courtis test is probably the only score of particular significance, and seems to indicate laziness and lack of interest. In view of the higher scores on the other tests this explanation may also hold for the low rating on definitions. On the whole the picture is that of a student with real ability who does not care to exert himself.

No. 16.

In spite of a good rating on the Thurstone test, this record indicates an individual of somewhat less than average ability. Although the academic rating is fair, the competency rating and the composite test rating are both low. Ratings below the middle quintile are found for the Taylor number test, the digit and syllable spans, the Trabue and definitions tests, while only the ratings for the Ausfrage, Courtis and memory tests are better than the average. It seems likely that this student supplements good retentiveness with more than the usual degree of industry in passing his college work.

No. 17.

Thurstone and competency ratings in the lowest quintile combined with the lowest composite test rating of the group indicate decidedly inferior ability in this case. Eight of the separate test ratings are below the middle quintile and only three are above. Low ratings on the Thurstone, Taylor, Trabue, Courtis, opposites and cylinder tests, all of which involve a definite speed factor, suggest that a slow rate of discharge is primarily responsible for the poor test performances of this individual. High ratings in the digit span, description, and definitions tests, in all of which the time element is relatively unimportant, seem to bear out this conclusion. An observation of the scores of the three memory span tests shows that as the material becomes more complicated the rating is lower. This man evidently needs time to think, and does well when the time is not limited. This fact explains the lack of correlation between the test ratings and the academic record, which is at least average, and it also emphasizes the undue weight given to the time factor in most mental tests.

No. 18.

This record displays the interesting combination of a very high academic rating with mediocre performance in the various mental tests. The record is quite comparable with that of student No. 1 with the exception that in this case nine of the thirteen test results are found in the middle quintile. High ratings in the Thurstone and Courtis tests suggest alertness, and this ability, in conjunction with a good rating on memory, may be partly responsible for the success in college work. It seems probable, however, that some motivation factor which cannot be measured by the test results has played an important part in the academic attainments of this student.

No. 19.

In this case the record, with the exception of the grades in psychology, is consistently above the average. Low ratings on the Courtis and cylinder tests might suggest a slow rate of discharge were it not for a very high rating on the Thurstone test. High scores on the three memory span tests, the Trabue, definitions, and memory tests show associability, retentiveness, and language ability, which may be looked upon as essential factors in intellectual development. The low rating on the cylinders hardly seems significant in view of the other test results, although it may indicate a deficiency in mechanical as contrasted with mental ability.

No. 20.

This record provides an interesting comparison with that of student No. 19. Although the psychology grades, competency rating, and Thurstone rating are identical, this student has a somewhat lower academic rating and a correspondingly lower composite test rating. Even the ratings for the separate tests show similar tendencies, but the scores for the Courtis and cylinder tests are lower here than in the preceding case. The most significant difference between the two records is found in the very low memory rating of this student, which places him definitely in the mediocre group.

No. 21.

This record is one of the most consistent to be found in the group and places the student definitely in the fourth quintile. The academic rating is quite high, the Thurstone and competency ratings are both "4", and the composite test rating is well above the average. The separate test scores indicate little, since all but two of the ratings are in the middle and upper quintile. Although the first cylinder trial was slow, the second trial compensated for this deficiency. There is no comment to make on this case other than a desire that mental tests might always correlate so closely with academic standing.

No. 22

While this record is, on the whole, mediocre, the academic rating is somewhat higher than might be expected in view of the low Thurstone and competency ratings. The latter may possibly be accounted for by the poor intelligence displayed in both cylinder performances, while good ratings on the tests requiring language ability, and particularly on the memory test, provide a satisfactory explanation for the fair academic rating. From the test results it seems probable that this individual has to apply himself to his studies in order to do passing work.

No. 23.

The failure in Psychology 2 is the only discordant note in an otherwise mediocre record. The composite test rating and that for the Thurstone test are about average for the group, while the competency rating is in the fourth quintile. The separate test results do not seem significant except for a high rating in memory. The poor work in psychology must probably be accounted for by lack of interest or failure to study.

No. 24.

The record here is comparable with that of student No. 15 in that a high composite test rating is contrasted with a low academic rating. In this case, however, the discrepancy is even more marked. The test rating is exceeded by only four members of the group, while only two have poorer college records. The separate test results present no solution to the difficulty since the ratings are high with only one exception. The competency rating is "4". It seems probable that this man is not particularly interested in his college work and is expending most of his time and energy in some kind of outside activity.

No. 25.

In this instance the record is consistently mediocre. All four of the general ratings are either in the middle quintile or slightly below the group average. The failure in Psychology 1 is hardly to be accounted for by the separate test results, which display no definite tendency, and was probably due to lack of application, since the student was able to pass the second course.

No. 26.

The rather high academic rating in this case seems to contradict the low Thurstone and composite test ratings. The low digit span and the poor rating on the memory test indicate that this student must be a hard worker in order to have received such high grades for his college courses. Good trainability as displayed in the second trial with the cylinders may be a significant factor in his academic work.

No. 27.

In this case a very high score on the Thurstone test correlates well with a high composite test rating and a high academic rating. "Distinguished" grades in both courses in psychology also indicate general superiority. A poor performance on the second trial with the cylinders which resulted in a competency rating of only "3" is the only flaw in an otherwise excellent record. Eight of the thirteen

tests are rated above the middle quintile and indicate nothing more than an unusually high level of general intelligence.

No. 28.

This record offers an interesting comparison with that of student No. 27. The composite test ratings and the competency ratings are identical in the two cases, while the academic ratings are very nearly so. Both students received the highest grade in both courses in psychology. In this instance, however, the Thurstone score is mediocre, and the ratings for the Trabue and cylinder tests are in the second quintile. The ratings on those tests which stress language ability are generally higher than in the preceding case, while the memory spans are conspicuously lower. These facts indicate a relatively low intelligence coupled with a rather high intellectual development. On the whole, the student is decidedly superior to the majority of the group.

No. 29.

The record in this case must be considered consistently good although it can hardly be compared with either of the two preceding cases in general excellence. The academic rating shows a "G" average and the psychology grades rate the student even higher. While the Thurstone rating is "4", the rating on estimated competency is higher than that in either of the preceding records. This rating is not substantiated by the results of the separate tests, only four of which are found to be above the middle quintile. These seem to point to intelligence rather than to intellectual organization, although it would be unsafe to make any specific diagnosis.

No. 30.

This record displays a relatively high test rating and a Thurstone rating in the fourth quintile contrasted with an average academic rating and unsatisfactory grades in psychology. While the separate test scores indicate somewhat erratic performances, very high ratings on the memory and cylinder tests show that this individual has unusual ability in some directions. It seems probable that lack of interest or want of application is responsible for the deficiency in psychology.

No. 31.

The mediocre composite test rating in this case does not correlate with the generally high level of the other ratings, all of which are in the fourth quintile. Although the separate test results are distributed through the five quintiles, they show no definite tendency which might be considered explanatory. Possibly the high

degree of trainability displayed in the second cylinder trial is significant, but it seems likely that this student either did not take the tests seriously or that some strong motivation factor has entered into his college work.

No. 32.

This record presents as great a contradiction as is to be found in the whole group. While only two students have academic records which exceed the rating in this case, only three have lower composite test ratings. Moreover, the estimated competency rating is "2" and the Thurstone test rating "5". Only three students have better grades in the two psychology courses. In the separate tests, low ratings were received on the Taylor number, digit span, Trabue, differences, definitions, and second cylinder trial. Only the syllable span and memory tests were rated higher than the middle quintile, the latter receiving the only "5" of the series. It seems hardly possible to explain the excellent academic record on the basis of good memory alone, and the only conclusion which can be reached is that the test results do not reflect the evident competency of this student.

No. 33.

All things taken into consideration, this is the poorest record in the group. The academic rating is low and one of the courses in psychology was not passed. The competency rating and that on the Thurstone test are both in the lowest quintile, the score on the latter test being the lowest made by any of the fifty students. The composite test rating is one of the lowest in the group, and only two of the separate test results are placed above the middle quintile. A rating of "5" in the memory test suggests that this ability may have enabled the student to stay in college. Low ratings on the Taylor, digit span, syllable span, Trabue, differences, opposites and definitions tests and the first trial with the cylinders indicate a very general deficiency. The test results in this case are quite similar to those in the record of student No. 32, but seem here to be really significant.

No. 34.

In this instance, the various ratings of the record correlate well to show better than average competency. The academic rating is good, the psychology grades very good, and the competency rating is in the fourth quintile. The Thurstone score is high, and while the composite test rating is only fair, the separate test results show no marked deficiencies. Low ratings in the Ausfrage and Taylor tests are not particularly significant, while higher ratings in the digit span, Courtis, and memory tests and second cylinder trial indicate asso-

ciability, speed, retentiveness and trainability. On the whole, the record shows no contradictions.

No. 35.

This record is consistent in so far as the composite test rating, the Thurstone rating and the competency rating are concerned. The test rating is equaled only by student No. 2, and both of the other ratings place this student in the highest quintile. In academic work, however, only an average rating is to be found, and the explanation must probably be based on lack of interest in studies or absorption in other activities. High ratings on the Taylor number, digit span, Trabue, Courtis, definitions and memory tests, on both trials with the cylinders, and on the Thurstone test indicate that this student has the ability to do excellent college work if he so desires.

No. 36.

Although a number of the separate test results are missing in this record, the ratings on the Thurstone test and estimated competency as well as the composite test rating indicate a rather high level of mentality. The academic rating, however, is one of the lowest in the group and shows that conditions and failures were received in a number of courses, even though the work in psychology was somewhat above the average. The evidence seems fairly conclusive that this man could do better college work if he wished to apply himself. Interest in outside activities probably explains the discrepancy between the test ratings and the academic record.

No. 37.

With the exception of a low rating on the Thurstone test, this record is consistently mediocre. The academic rating, competency rating and composite test rating all appear in the middle quintile. A good performance on the first trial with the cylinders, followed by an excellent second trial, indicate intelligence and trainability, while a low rating on the memory test may explain the mediocre college record.

No. 38.

The record in this instance is consistently below the average for the group and may be considered typical of the second quintile. The academic rating is low, the psychology grades merely passing, the competency rating and the Thurstone rating are both "2", and the composite test rating decidedly below the average. Low ratings were received on the Taylor number, idea span, description, and Trabue tests, while the digit span, definitions and cylinder tests were rated above the middle quintile. No ratings in the highest

quintile appear. An analysis of these results seems to indicate good associability and intelligence coupled with rather deficient intellectual organization. This man would probably be more successful in business than in an academic or professional vocation.

No. 39.

This record disputes with that of student No. 33 the distinction of being the poorest in the group. The fact that the student was excluded from the University at the end of the session gives peculiar interest to this case. An observation of the grades received in college courses discloses the significant fact that eight units of work were assigned a grade of "D", while an equal number received a "G". Eight units of credit were merely "Passed", conditions were given for three units, and the remaining eight units received the grade "F". Passing grades were assigned for both courses in psychology. This unusual distribution of grades suggests specific ability along certain lines with marked variations in interest. The student would probably have received "Distinguished" grades in all of his college work if he had been allowed free election of courses. Low ratings on the Thurstone and Courtis tests show that he cannot think quickly, while poor scores in the Trabue and memory tests indicate deficiency in imagination and retentiveness. High ratings on the Taylor number and cylinder tests show that there is no deficiency in the rate of discharge of energy, and that distribution of attention and intelligence are both above the average. It seems probable that this man, now being free to follow his own inclinations, will be successful in the vocation which he chooses. The case is particularly interesting as an example of the influence of special abilities and of motivation in the behavior of the individual.

No. 40.

In this case a very low competency rating is contradicted by a composite test rating only slightly below the average and Thurstone and academic ratings in the fourth quintile. The competency rating was doubtless influenced by very poor performances in both cylinder trials, but this deficiency in intelligence is compensated for by high ratings in the syllable and idea spans, Courtis, definitions and memory tests. In other words this student has the associability, alertness, language ability and retentiveness necessary to do good college work. It is possible, also, that lack of interest in the tests may have affected the significance of the results.

No. 41.

This is a consistently mediocre record with the exception of the psychology grades, which are slightly above the average, and

the competency rating, which is very high. The academic rating is slightly below the median and the composite test rating is median for the group. The Thurstone score is placed in the middle quintile. High ratings are shown for the Ausfrage, Courtis, and first cylinder trial. The latter, however, is offset by a very poor performance in the second trial with the cylinders. Low ratings also appear for the Taylor number, digit and syllable spans, Trabue, and memory tests. These results indicate rather poor general intelligence and suggest that the competency rating is too high.

No. 42.

Every one of the principal ratings in this record occurs in the highest quintile, and the student must be ranked definitely with the leaders of the group. High ratings on the Ausfrage, description, Trabue, and memory tests and on both cylinder trials show good observation, imagination, retentiveness and intelligence. A low rating on the digit span is neutralized by a high idea span. Other low ratings on the Courtis and opposites tests do not seem significant. On the whole the record is unusually consistent and justifies the high competency rating.

No. 43.

Although the composite test rating in this case is about average for the group, the academic rating is decidedly inferior. The competency rating is the lowest given to any member of the class, and is based on very poor performances with the cylinders. Although this student seems to lack intelligence, high ratings were obtained in the Ausfrage, Trabue, and definitions tests. Low ratings for the Taylor number, digit span, and Courtis tests indicate a consistently poor performance in those tests which do not involve language ability. The good ratings in the strictly intellectual tests suggest that outside activities are responsible for the low academic rating.

No. 44.

This record seems to be typical of the fourth quintile. The academic record shows a preponderance of "Good" grades, and this mark was received for both courses in psychology. The competency rating is "4" and the Thurstone rating "3". The composite test rating is one of the best in the group, although fifth quintile ratings appear only for the Ausfrage test and the first cylinder trial. Other test ratings show a high level of general intelligence with no significant disabilities.

No. 45.

An academic record in the fourth quintile is accompanied in this case by a Thurstone score in the middle quintile, a competency rating in the second, and a composite test rating in the lowest quintile of the group. This unanimous absence of correlation is also shown in the separate test results where ratings in all five quintiles appear. A high rating on the Taylor number test suggests good distribution of attention, but even this ability must have been lacking in the cylinder performances. The test results show no definite tendency, but display a low level of general intelligence. The high academic rating notwithstanding, this student falls below the middle quintile of the group in competency.

No. 46.

The lowest academic rating in the group is displayed by this senior, who, nevertheless, was able to graduate with his class. While the Thurstone score is poor, the competency rating and the composite test rating are both high. The separate test results are low for digit and idea spans, but high for most of the other tests with exceptionally good performances on the cylinder test. This student was evidently doing no more college work than was necessary to obtain his degree, and was probably interested in outside activities.

No. 47.

The competency rating, composite test rating, and academic rating agree in placing this student in the second quintile. The rating on the Thurstone test is very high, and the grades in psychology the poorest in the group, consisting of an "N" for the first course and a "Failure" for the second. High ratings on the Thurstone and Courtis suggest a rather quick mind when familiar operations are involved, while the very low ratings on the cylinder test indicate inability to meet a new problem successfully. Since the subject-matter of the courses in psychology is quite unlike that of most college courses, the inability of the student to adapt himself to the new situation is probably the cause of his deficient work in this subject. Although the result for the memory test is missing, a high rating in that ability may be predicted.

No. 48.

In this record the composite test rating, the competency rating and the Thurstone rating indicate a very high level of general intelligence. The academic rating, however, is far below the average for the group. Of the separate test results, only two fall below the middle

quintile. The low ratings on the Trabue and Courtis tests are difficult to explain in the light of the other test ratings, five of which are in the highest quintile. Excellent associability, language ability, retentiveness, and intelligence are displayed in the various test scores, and the only explanation of the relatively poor college grades seems to lie in lack of interest or absorption in outside activities.

No. 49.

Although the Thurstone score, the competency rating, the composite test rating, and the grades in psychology agree in placing this student below the middle quintile, the academic rating is the median for the group. As is frequently the case where this situation is encountered, the rating on the memory test is high. In addition to this test only the Ausfrage and the syllable span were rated higher than the middle quintile, while eight of the thirteen tests fell below that level. It seems certain that more than the usual amount of industry is expended by this individual on his college work.

No. 50.

This record is quite similar to that of student No. 49 with the exception that the composite test rating is slightly lower and the academic rating somewhat higher than in the preceding case. Here, however, the Thurstone rating is high and the competency rating and psychology grades average. Of the separate test results, only the rating on the memory test is in the highest quintile. The ratings for the Taylor number, description, Trabue, Courtis and first cylinder tests are in the lowest quintile. The second trial with the cylinders indicates good trainability, which with the assistance of an unusually good memory may account for the high academic rating. On the other hand, lack of effort in the tests may be responsible for the low composite test rating, and is suggested by the high score on the Thurstone test.

Summary.

A scrutiny of the analyses of the fifty individual records shows that these may be separated into two general groups. In twenty-six cases the correlation between the various ratings is close enough to present fairly conclusive evidence of the relative performance level of the student. These cases, in turn, naturally fall into five classes corresponding roughly with the points of a five-division scale, which may be referred to here as very good, good, medium, poor, and very poor. Seven records are so consistently high as to warrant a place in the first group, while five more are distinctly better than the average and may be considered "good". Eight cases occur in the

"medium" class, and of the six which fall below this level two are "poor" and four show such a general inferiority as to justify placement in the lowest group. The twenty-four remaining records, which display a decided lack of correlation between the various major ratings, exhibit two opposing tendencies. In fourteen cases the academic rating is higher than would be predicted from the test results, while in the ten remaining cases the Thurstone score, competency rating and composite test rating would seem to indicate better scholastic ability than is displayed in the academic rating and psychology grades. The following summary shows the classification of each individual record.

Classification of Individual Records.

I. Cases showing general correlation of ratings:

Very good.....2, 13, 14, 27, 28, 42, 44
Good.....11, 18, 19, 21, 29
Medium.....4, 5, 6, 20, 23, 37, 40, 41
Poor.....22, 47
Very poor.....3, 33, 38, 39

II. Cases where correlation is lacking:

High academic, medium mental.....1, 9, 12, 31, 34
High academic, low mental.....26, 32, 45
Medium academic, low mental.....8, 10, 16, 17, 49, 50
High mental, medium academic.....15, 30, 35
High mental, low academic.....24, 36, 46, 48
Medium mental, low academic.....7. 25, 43

Although in some cases the evidence is not so clear cut as the summary above may seem to indicate, the classification nevertheless is justified by the data at hand. It also seems reasonable to attribute the absence of correlation shown in the second group of records to variations in motivation and other external factors which have not as yet lent themselves to quantitative measurement. Of two men who have the same composite test rating and who may be assumed to possess equal competency, one may be intensely interested in his studies and impelled by a consuming ambition to gain the greatest possible benefit from his college course, while the other is content to do only the amount of work necessary to fulfil the minimum scholastic requirements and seeks to excel in athletic or social activities. Again, the first student may be devoting all of his time and effort to college work, while the second is compelled to expend much of his energy

in supporting himself. Certainly no series of mental tests will correlate closely with academic standing until some satisfactory method of evaluating these factors external to competency has been devised. At present it is possible to do no more than call attention to the lack of correlation and attempt to explain the discrepancies in the most logical manner.

Conclusions.

(1) The psychologist should engage in the analysis and evaluation of the "ability" components of the college student's competency rather than in the correlation of general intelligence tests with academic grades.

(2) The abilities required for scholastic success, under the present methods of college instruction and grading, are not all of the abilities comprising individual competency. Hence the failure of test results to correlate with college grades. The better the general intelligence test, the smaller will be the correlation with academic standing.

(3) College grades will provide more satisfactory material for statistical treatment when each institution adopts a standard distribution of grades and provides for supervision by some administrative officer.

(4) Tests for college students must be devised which place less dependence upon time measurement, which have a higher reliability coefficient, and which are of greater difficulty, than most of the tests now available.

(5) Motivation and environmental and economic conditions have not as yet yielded to quantitative treatment. Until they do, it will not be possible to predict with accuracy the success of a student in college or in any other field of endeavor.

(6) Test ratings such as those presented here should be made available to deans, faculty advisers, and committees dealing with scholastic deficiency. In many instances this information would be of value to the student, also, providing him with educational or vocational guidance.

(7) A "follow up" of the fifty students who have provided the material for this study will be published at some future date.

(8) Only after many investigations are at hand with diagnoses carefully followed up over a period of years will psychological diagnosis and orthogenic guidance become as reliable for the normal individual as it is now for the subnormal.

BIBLIOGRAPHY.

1. Wissler, Clark. The Correlation of Mental and Physical Tests. *Psychological Review Monograph Supplement 3*, 1901, No. 6, 1–61.

2. Calfee, M. College Freshmen and Four General Intelligence Tests. *Journal of Educational Psychology*, 1913, 4, 223–231.

3. Rowland, E., and Lowden, G. Report of Psychological Tests at Reed College. *Journal of Experimental Psychology*, 1916, 1, 211–217.

4. Waugh, Karl T. A New Mental Diagnosis of the College Student. *New York Times Magazine Supplement*, January 2, 1916.

5. Kitson, H. D. Scientific Study of the College Student. *Psychological Monograph No. 98*, 1917. Pp. 81.

6. King, I., and McCrory, J. Freshmen Tests at the State University of Iowa. *Journal of Educational Psychology*, 1918, 9, 32–46.

7. Caldwell, H. H. Adult Tests of the Stanford Revision Applied to College Students. *Journal of Educational Psychology*, 1919, 10, 477–488.

8. Rogers, A. L. Mental Tests as a Means of Selecting and Classifying College Students. *Journal of Educational Psychology*, 1920, 4, 181–192.

9. Humpstone, H. J. Some Aspects of the Memory Span Test. Experimental Studies in Psychology and Pedagogy, 7. Psychological Clinic Press, Philadelphia, 1917. Pp. 31.

10. Terman, L. M. The Measurement of Intelligence. Houghton-Mifflin Company, Cambridge, 1916. Pp. 362.

11. Trabue, M. R. Completion-Test Language Scales, Teachers College, Columbia University, 1916. Pp. 118.

12. Courtis, S. A. The Courtis Standard Tests. Department of Co-operative Research, Detroit, 1914. Pp. 125.

13. Whipple, G. M. Manual of Mental and Physical Tests, Part II. Warwick and York, Baltimore, 1915. Pp. 336.

14. Paschal, F. C. The Witmer Cylinder Test. The Hershey Press, Hershey, Pa., 1918. Pp. 54.

15. Finkelstein, I. E. The Marking System in Theory and Practice. Warwick and York, Baltimore, 1913. Pp. 87.